SCIENCE LIBRARY

JAN 31 1931

Vol. XIV, No. 1 *PSYCHOLOGICAL REVIEW PUBLICATIONS* FEBRUARY, 1931

18

# Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
RAYMOND DODGE, YALE UNIVERSITY (*Monographs*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY, Business Editor

## CONTENTS

*The Refractory Phase of Voluntary and Associative Responses:* C. W. TELFORD 1

*The Relation of Memory to Intelligence:* E. B. BOLTON..... 37

*An Experimental Test of the Law of Assimilation:* K. S. YUM..... 68

*The Factors of Speed and Power in Tests of Intelligence:* F. S. FREEMAN. 83

*The Variability of Affective Judgments upon Odors:* J. G. BEEBE-CENTER 91

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

# PUBLICATIONS

OF THE

# AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
RAYMOND DODGE, YALE UNIVERSITY (*Monographs*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)
S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*J. of Exp. Psych.*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index and Abstracts*)
HENRY T. MOORE, SKIDMORE COLLEGE (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY, Business Editor

WITH THE CO-OPERATION OF
MANY DISTINGUISHED PSYCHOLOGISTS

## PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

## PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

## JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 500 pages.

## PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

## PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

## PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

## JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

## ANNUAL SUBSCRIPTION RATES

**Review:** \$5.50 (Foreign, \$5.75). **Index:** \$4.00 per volume.
**Journal:** \$6.00 (Foreign, \$6.25). **Monographs:** \$6.00 per volume (Foreign, \$6.30).
**Bulletin:** \$6.00 (Foreign, \$6.25). **Abstracts:** \$6.00 (Foreign, \$6.25).
**Abnormal and Social:** \$5.00 (Foreign, \$5.25). Single copies \$1.50.
Current numbers: **Review or Journal, \$1.00; Abstracts, 75c; Bulletin, 60c.**

## COMBINATION RATES

**Review and Bulletin:** \$10.00 (Foreign, \$10.50).
**Review and J. Exp.:** \$10.00 (Foreign, \$10.50).
**Bulletin and J. Exp.:** \$11.00 (Foreign, \$11.50).
**Review, Bulletin, and J. Exp.:** \$15.00 (Foreign, \$15.75).
**Review, Bulletin, J. Exp., and Index:** \$18.00 (Foreign, \$18.75).

---

Subscriptions, orders, and business communications should be sent to the

## PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, N. J.

Science
Wahr

# CONTENTS OF VOLUME XIV


BASS, M. J. **Differentiation of the Hypnotic Trance from Normal Sleep** ..... 382
BEEBE-CENTER, J. G. **The Variability of Affective Judgments upon Odors** ..... 91
BOLTON, E. B. **The Relation of Memory to Intelligence** ..... 37
CARROLL, R. P. **Practice in Rating** ..... 299
COPELAND, H. A. **An Apparatus for Recording Electrical Changes** ..... 180
COREY, S. M. **An Experimental Study of Retention in the White Rat** ..... 252
CROSLAND, H. R. **Letter-Position Effects, in the Range of Apprehension Experiment, as Affected by the Number of Letters in Each Exposure** ..... 477
CUFF, N. B. **A Study of Eyedness and Handedness** ..... 164
DIMMICK, G. B. **An Automatic Rotary Switch for Use with the Ranschburg Exposure Apparatus for Continuous Multiple Choice Work** ..... 303
FREEMAN, F. S. **The Factors of Speed and Power in Tests of Intelligence** ..... 83
FREEMAN, G. L. AND LINDLEY, S. B. **Two Neuro-Muscular Indices of Mental Fatigue** ..... 566
FORD, A. **The Correlator** ..... 155
HELSON, H. AND KING, S. M. **The Tau Effect: An Example of Psychological Relativity** ..... 202
HERREN, R. Y. **The Effect of Stuttering on Voluntary Movement** ..... 289
HUSBAND, R. W. AND LUDDEN, M. J. **Sex Differences in Motor Skills** ..... 414
JERSILD, A. **Memory for the Pleasant as Compared with the Unpleasant** ..... 284
JONES, H. E. **Dextrality as a Function of Age** ..... 125
KARWOSKI, T. **The Memory Value of Size** ..... 539
KRUEGER, R. G. **An Improved System of Kymographic Recording** ..... 176
KRUEGER, R. G. **The Influence of Repetition and Disuse upon Rate of Hypnotization** ..... 260
MCGEOCH, J. A. **The Influence of Four Different Interpolated Activities upon Retention** ..... 400
MCGEOCH, G. O. **The Intelligence Quotient as a Factor in the Whole-Part Problem** ..... 333
MCTEER, W. **A New Device for Serial Order Learning** ..... 446
MCTEER, W. **A Study of Certain Features of Punishment in Serial Learning** ..... 453
MILES, W. R. **Elevation of the Eye-Balls in Winking** ..... 311
MOSS, F. A., ROE, J. H., HUNTER, O. B., FRENCH, L. AND HUNT, T. **The Measurement of Fatigue by Physiological Methods** ..... 423
O'SHEA, H. E. **Errors of Measurement in a Test and a Retest** ..... 439
PERKINS, F. T. **A Further Study of Configurational Learning in the Goldfish** ..... 508
PETERSON, J. AND ALLISON, L. W. **Controls of the Eye-Wink Mechanism** ..... 144
RAZRAN, H. S., AND WAGNER, M. E. **The Summation Method in Statistics** ..... 270
REED, H. J. **The Influence of a Change of Conditions upon the Amount Recalled** ..... 632
SEASHORE, S. **The Aptitude Hypothesis in Motor Skills** ..... 555
SMITH, M. **Group Judgments in the Field of Personality Traits** ..... 562
STROUD, J. B. **Apparatus for Measuring Muscular Tensions** ..... 184
STROUD, J. B. **The Rôle of Muscular Tensions in Stylus Maze Learning** ..... 606

TELFORD, C. W. **The Refractory Phase of Voluntary and Associative Responses**.. 1
THURSTONE, L. L. **Rank Order as a Psychophysical Method**..... 187
TRAVIS, L. E. AND LINDSLEY, D. B. **The Relation of Frequency and Extent of Action Currents to Intensity of Muscular Contraction**..... 359
VICKERY, K. **The Effects of Change of Work on the Work Decrement**..... 218
WINSOR, A. L. **Some Quantitative Characteristics of Parotid Secretions**..... 242
YOUNG, P. T. **The Rôle of Head Movements in Auditory Localization**..... 95
YUM, K. S. **An Experimental Test of the Law of Assimilation**..... 68

Journal of
Experimental Psychology

VOL. XIV, No. 1 FEBRUARY, 1931

# THE REFRACTORY PHASE OF VOLUNTARY AND ASSOCIATIVE RESPONSES [1]

BY C. W. TELFORD, PH.D.

*University of North Dakota*

CONTENTS

I

PAGE

Introduction ..... 1
Historical sketch ..... 3
Statement of the problem ..... 6

II

Experimental ..... 7
Reaction time as influenced by the length of the interval between stimuli .... 7
Accuracy in judging linear magnitudes as influenced by the length of the interval between judgments ..... 12
Repetitive tendencies in an associative process as influenced by the length of interval between responses ..... 21

III

Interpretations and conclusions ..... 27
Practical applications ..... 33
Suggestions for further study ..... 34

## I. INTRODUCTION

"Restricted in the method of direct experimentation to techniques that may be applied to intact and complex nervous systems, the experimental psychologist is peculiarly dependent on the results of physiological experiments with isolated tissues and simple systems both for his working hypotheses

[1] The writer is greatly indebted to Dr. Joseph Peterson for suggestions and criticisms in connection with the present study.

and his elementary facts. But he must take scientific responsibility for testing the extension of the hypotheses and facts to the complex conditions of human experience and personality." Thus wrote Dodge (1926, p. 3). In the present study the writer has attempted to follow the lead of Dodge in ascertaining how far the post-stimulation phenomena, as discovered in isolated tissues and simple reflexes, are found in complex motor and associative responses and in determining what relation these phenomena or effects may have to the more complex problems of behavior.

Two of the time-honored traditions in psychology are stated in the alleged laws of frequency and recency. According to these laws whenever a nervous process traverses a certain pathway in the nervous system it leaves the resistances in that pathway diminished or less than it found them. According to the latter law the more recently the passage of the impulse the less the resistance, and according to the former the greater the number of times the neural process operates the less the resistance. Of late considerable evidence has been amassed which shows that these laws have some very serious limitations. Peterson (1917), Kuo (1922), and of late, Thorndike (1927*a*), and Dunlap (1928), have all presented facts which appear to be unexplainable by such hypotheses.

It has also been generally recognized that the irritability of a living system undergoes diminution as the result of long-continued stimulation. Within recent times it has been shown that a single momentary excitation is likewise followed by such a reduction of specific irritability. This period of decreased excitability subsequent to stimulation is known as the refractory period. The period during which the re-excitation is not possible by maximal stimuli is called the absolute refractory period. The period during which a response is evoked only by abnormally intense stimuli,—or during which the response to normal stimuli is abnormally weak,—is known as the relative refractory period. Such a condition seems to be a wide-spread consequence of the reaction of irritable tissue. It was first discovered in heart

muscle and has since been demonstrated in the responses of nerve fibers and skeletal muscles as well as in a number of reflexes. In fact this effect of the reaction of irritable tissue seems to have been found whenever it has been sought with adequate technique.

The period is a few thousandths of a second in length in the nerve fiber. In some reflexes it may last for several seconds. It has been shown that it may be prolonged indefinitely in complex neural paths by stimuli which succeed one another so rapidly that the path fails to recover normal excitability (Lucas and Adrian, 1917). Fröhlich and Verworn (1913), originated the hypothesis that inhibition is due to the development of a prolonged refractory phase in the tissues concerned. This has been supported by a long series of brilliant experimental investigations by Lucas and Adrian (1917) in England, and Forbes (Forbes, Querido, Whitaker, Hurxthal, 1928) in America.

Since learning always involves a negative factor, *i.e.* the elimination of unsuccessful responses, it is possible that it may involve the inhibition of useless responses by the development of a prolonged refractoriness as suggested by Dodge (1927, 1927*a*).

*Historical sketch.*—The refractory phase was first discovered in cardiac tissue. There it is of long duration and needs no elaborate timing apparatus for its measurement. The classical experiment on the refractory phase of heart muscle is that of Marey (1876). He worked with the frog's heart and published curves showing that the irritability of the heart in response to artificial stimulation was greatly reduced during systole, and that recovery took place during diastole. Marey called this period of reduced irritability 'phase réfractaire.'

McWilliam (1888), Englemann (1895), Schultz (1906), and more recently Hermann and Umrath (1927) have determined more exactly the duration of the refractory period in cardiac tissue. These workers have studied the hearts of frogs, cats, dogs, rabbits, hedgehogs, guinea-pigs and rats. They worked on hearts in situ as well as on excised hearts.

Their results agree in general, and indicate that the refractory period is limited to the time of systole plus a portion of diastole. That portion of diastole during which the tissue is refractory but can be forced to respond by stimuli sufficiently strong is known as the 'variable' or 'relative' refractory period, whereas the time of systole during which the tissue does not respond even to maximal stimulation is called the 'absolute' refractory period.

For several years the refractory phase was thought to be a characteristic of cardiac tissue alone. However, some twenty years after Marey's investigation Broca and Richet (1897) found an analogous refractory period in the tissues of the motor centers of the cerebral cortex of the dog. They found that in a normal dog the greatest number of contractions that could be brought about by electrical stimulation was ten per second, giving a refractory period of more than .1 second. This period of nonirritability was found to be prolonged to almost ten times its normal length in a dog affected with chorea.

Gotch (1910) and his students (Gotch and Burch, 1899; Boycott, 1899; Bazett, 1908) were apparently the first to demonstrate the existence of a refractory phase in nerve and skeletal muscle. Lucas in (1909) published the first of a long series of excellent researches on the same subject. Verworn and his school (Verworn, 1913), working along the same lines, have shown, with much plausibility, that the refractory period is an elementary phenomenon of fatigue.

Adrian and Lucas divide the phenomenon of recovery into three distinct periods. There is first an 'absolute refractory period' during which the tissue is inexcitable. Then follows the 'relative refractory period' during which the excitability returns gradually to normal. This period of impaired excitability is followed by one in which the irritability is greater than normal. This they call the 'supernormal phase.'[2] Then comes a return to normal (Adrian and Lucas, 1912; Lucas and Adrian, 1917).

[2] Alexander Forbes in *Foundations of Experimental Psychology* (Carl Murchison, ed.), 1929, 139–140, has the following to say on this same topic: "Another feature common to muscle and nerve, and of the greatest importance, is the *refractory phase.*

A refractory phase has also been found in various reflexes. It was observed in the eyelid reflex by Zwaardemaker and Lans (1899) and by Dodge (1913 and 1927); in the knee jerk by Dodge (1911 and 1927), Denny-Brown and Liddell (1927), and Strughold (1927); in the swallowing reflex of the cat by Zwaardemaker (1904); in the flexion reflex by Forbes, Querido, Whitaker and Hurxthal (1928), and by Sherrington (1906); in the 'extensor thrust' by Sherrington (1906); and in the start reflex of the guinea-pig by Dodge and Louttit and also by Cohen (1929).

As would be expected from the all-or-none law, the strength of the stimulus has no effect on the length of the period of diminished excitability following response (Hermann and Umrath, 1927). That raising the temperature decreases, and lowering the temperature increases the length of the period has been shown by Gotch and Burch (1899), Gotch (1910), Boycott (1899) and Verworn (1913).

Verworn showed that by gradually depriving the system of oxygen the length of the refractory period can be increasingly prolonged until finally irritability does not reappear at all. Alcohol and potassium salts lengthen, while calcium salts shorten the refractory period (Schultz, 1906).

Dodge (1927) reports that intentional reinforcement of quadriceps contraction in the knee-jerk of humans obliterates all evidence of the relative refractory phase. That is, both first and second of the paired reflexes may become maximal contractions by voluntary reinforcement. Further details are not given in the report. Strughold (1927), studying the same reflex in humans, found that facilitation by voluntary contraction of the extensors in opposing a weight hung over

When a propagated disturbance is initiated it sweeps over the tissue, leaving it refractory to further stimulation for a brief period of time. Following the absolute refractory phase when the tissue cannot respond, there is a period of recovery, the relative refractory phase, during which the threshold of excitation is abnormally high, and the size of the response that can be evoked is subnormal, both returning gradually to normal. Adrian has shown that in the case of a nerve immersed in a slightly acid fluid there is, following the relative refractory phase, an apparently *supernormal* phase in the recovery of excitability. He further showed that the excitability becomes no greater than would be found if the tissue were in neutral fluid. The supernormality is only relative, due to the fact that in an acid medium the final resting stage is subnormal when compared to the condition of the tissue at neutrality."

a pulley reduces the resulting refractory period from three seconds to less than one second. Inhibition of the quadriceps by contraction of the flexors prolongs the period to six seconds. These findings may have considerable significance and should be verified and extended by further study.

*Statement of the problem.*—These studies indicate that the refractory phase is a universal post-stimulation phenomenon of sensitive tissues. So far, no standard technique has been developed for investigating the refractory period of voluntary processes. It is not known whether such a phenomenon exists and plays a part in complex behavior. Dodge (1927 and 1927*a*) has pointed out certain facts, based principally on general observation, which seem to indicate that such a tendency does operate in voluntary responses.

If such a period of decreased excitability does follow voluntary response, it may possibly show itself (1) as a more or less pronounced tendency not to repeat a given act, (2) as an increase in time of response to stimuli occurring at less than the period necessary for recovery, or (3) as a decrease in accuracy of responses which occur very rapidly, or in combinations of these phenomena.

That the first might be true seems evident from the very nature of the refractory period, with its decreased excitability. That an abnormal delay in response to stimuli occurring more often than a critical frequency may be expected, seems evident from the work of Gotch (1910), Lucas and Adrian (1917), Dodge (1927), Forbes, Ray and Griffith (1923), and Gasser and Erlanger (1925). These investigators all found that the latency of the response to a stimulus applied during the refractory period is of abnormal length. The increase in delay is augmented in proportion as the succeeding stimuli approach the period of complete inexcitability. Lucas (1911) suggests that by the delay of the response to a second stimulus one can determine the duration of a modified state which is left in the wake of a previous response. This is essentially one of the methods used in the present investigation. [3]

[3] For further discussion of the nature and cause of this delay, interested readers are referred to the excellent reviews by Forbes, Ray and Griffith (1923), and Gasser and Erlanger (1925).

On a purely *a priori* basis it seems probable that successive judgments made at short intervals might be less accurate than those occurring less frequently. At least, it seems worth investigating.

The present study was planned, in view of such possibilities, to determine the effect of the length of the interval between stimuli (1) on simple reaction time, (2) on accuracy of judgment and (3) on tendencies to repeat. Stated briefly our problem is this: Do voluntary responses, judgments and simple associative processes produce effects in the organism which serve as a barrier against immediate repetition, that can be identified with a refractory phase? If so, is the subsequent recovery and return to normal comparable to that found in isolated tissues and reflexes?

## II. Experimental

Our experimental work is divided into three parts. The first determines in what way the speed of simple reactions to auditory stimuli is influenced by the length of the interval between stimuli. The second investigates the relationship between accuracy in judging the lengths of lines and the time intervening between the presentations of the lines. The third is an attempt to determine the influence of the length of the interval between stimuli on tendencies to repeat in associative responses.

*Reaction time as influenced by the length of the interval between stimuli.*—Since the beginning of the reaction time experiments a voluminous literature has accumulated on the subject, but as none of the previous studies relates directly to our problem, we shall not review them here. Readers interested in this history are referred to Johnson (1923) and Gatewood (1920).

Our method consisted essentially in recording serial reaction times to simple auditory stimuli. The stimuli were given at $\frac{1}{2}$, 1, 2, and 4 seconds intervals according to a prearranged chance order. These intervals were selected on the basis of preliminary experimentation on three subjects, the results of which showed that at intervals less than $\frac{1}{2}$ second

the second stimulus was often received before the subject had responded to the one preceding. The average reaction times to stimuli occurring at 8- and 16-seconds intervals were practically the same as the time at the 4-seconds intervals. Consequently these longer intervals were discarded and the four mentioned were used in the main series of experiments.

The reaction times were recorded on a belt of smoked paper revolving on a kymograph and timed by a simultaneous graph of a 50 d.v. electrically driven tuning fork. Stimuli were administered by means of an electrically operated sounder, controlled by the experimenter. In the same circuit with the sounder were placed an electric marker which recorded on the kymograph record the exact time of giving the stimulus, a metronome making contacts every half second, and a telegraph key operated by the experimenter.

The subject responded by pressing down on a telegraph key which closed a second circuit. In this circuit was a second electric marker which recorded the exact time of response. The speed of reaction was determined by counting for each reaction the number of tuning fork vibrations between the record of the stimulus and that of the response.

The reacting subject, together with the sounder and key for response, was in a separate room far enough removed from the recording apparatus not to be disturbed by the noise of the kymograph and of the muffled metronome. The sounder was also enclosed in a box to prevent the subject from watching it and thus responding to a visual rather than to the auditory stimulus.

The subjects consisted of 29 graduate students (15 men and 14 women) taking work in psychology and educational psychology in the Jesup Psychological Laboratory. Only two of these subjects had had any previous experience in reaction time experiments.

Each subject was seated before the table upon which the sounder and telegraph key were fastened. He was instructed to "respond as quickly as possible by pressing down the telegraph key every time the sound is heard." The subject

was informed that the stimuli would be given at irregular intervals and that he was to be ready to respond at any time. The stimulus was in all cases given a few times before the experimental series began, in order to familiarize the subject with the sound and to make sure that he understood just what was to be done.

The experimenter with the order of intervals before him started the metronome and kymograph; then he pressed down the stimulus key so that the metronome on its next stroke would make contact on the particular interval at which the stimulus was to be given. When the metronome had made contact the experimenter released his key, unless the next interval happened to be a half-second one. In such a case, the stimulus key was held down until the metronome made contact again. Thus the desired intervals were readily obtained after preliminary practice and small irregularities due to the experimenter's inability to give stimuli at exactly the correct intervals were avoided. The record showed clearly any errors made in giving the stimulus and all such records were discarded. There were very few of these.

Each of the four intervals occurred approximately 25 times. The 100 responses of each subject were all made at a single sitting and were given in two series of 50 each. A two minute rest period intervened between the two series.

*Results.*—The results of this experiment are given in Table I. An examination of this table shows that the ½-

TABLE I

COMPARISONS OF THE MEAN REACTION TIMES OF 29 SUBJECTS WITH VARYING TIME INTERVALS BETWEEN STIMULI

| Approximate intervals between stimuli in seconds | ½ | 1 | 2 | 4 | | |
|---|---|---|---|---|---|---|
| Exact interval between stimuli in $\sigma$ | 505.6 | 1011.2 | 2022.4 | 4044.8 | | |
| Number of responses | 697 | 693 | 734 | 866 | | |
| Mean reaction time in $\sigma$ | 335 | 241 | 245 | 276 | | |
| Standard deviation | 63.6 | 42.5 | 50.6 | 56.2 | Diff. ½ and 2 | Diff. ½ and 4 |
| Difference between successive means | | 94.0 | −4.0 | −31.0 | 90.0 | 59.0 |
| P.E. of diff. (Long formula) | | 6.0 | 3.6 | 5.8 | 9.5 | 12.1 |
| Difference/P.E. difference | | 15.67 | −1.11 | −5.34 | 9.47 | 4.88 |
| Difference/P.E. of shorter interval | | 2.19 | − .14 | − .96 | 2.09 | 1.37 |

second interval is the least favorable for rapid response. The average reaction time at this interval is 335 $\sigma$. The most favorable intervals are 1 and 2 seconds. The average time at these intervals is reduced to 241 $\sigma$ and 245 $\sigma$, respectively. As the interval is increased beyond two seconds there is a subsequent increase in reaction time. With a 4-seconds interval between successive stimuli the average reaction time is 276 $\sigma$. Row eight of Table I shows the difference between the reaction times at $\frac{1}{2}$-second and at 1-second intervals to be highly reliable statistically, the difference being 15.67 times its probable error. There is practically no difference between the 1-second and the 2-seconds intervals. The corresponding difference between the 2-seconds and the 4-seconds interval is reliable, being 5.34 times its probable error. The ninth row of the table gives the size of the difference in terms of the probable error of the shorter interval. These results are shown graphically in Fig. 1.

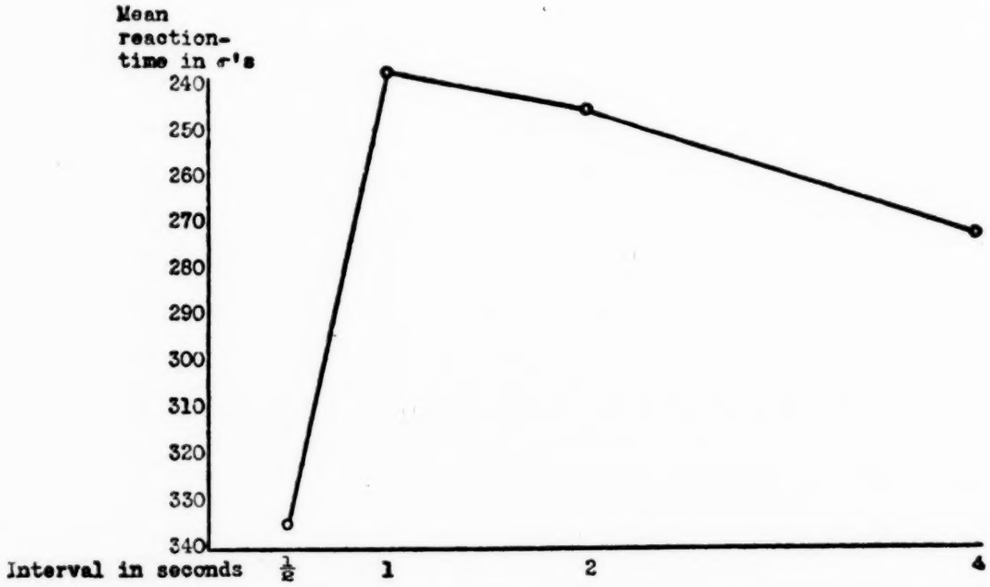



FIG. 1. Mean reaction-time with varying time-intervals between stimuli.

Verworm (1913) has shown that during fatigue the refractory phase is somewhat prolonged. Dodge (1927) and Lashley (1924) have suggested that learning may involve a change in the refractory periods of the systems involved.

To determine the effect of whatever practice and fatigue may have been involved in these reactions, a comparison was made between the means of the first and last halves of each subject's reaction times at the various intervals. These results are shown in Table II.

TABLE II

COMPARISON OF THE DIFFERENCES BETWEEN THE MEANS OF THE FIRST AND SECOND HALVES OF THE SERIES OF REACTION TIMES OF 29 SUBJECTS WITH VARYING TIME INTERVALS BETWEEN RESPONSES

| Interval in seconds | ½ | | 1 | | 2 | | 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First or second half..... | 1st | 2d | 1st | 2d | 1st | 2d | 1st | 2d | | | | |
| Mean reaction time in $\sigma$'s.... | 350 | 320 | 246 | 236 | 251 | 237 | 287 | 269 | | | | |
| Difference between 1st and 2d half of each series..... | | −30 | | −10 | | −14 | | −18 | | | | |
| P.E. of differences..... | | 7.01 | | 4.77 | | 5.14 | | 5.65 | | | | |
| Standard deviation of dist.... | 71.0 | 69.4 | 51.0 | 47.2 | 60.6 | 49.6 | 67.0 | 52.2 | (Diff. ½ and 2) | | (Diff. ½ and 4) | |
| Difference between successive means[1]... | | | 104 | 84 | 5 | 1 | 36 | 32 | 99 | 83 | 63 | 51 |
| P.E. of difference between successive means..... | | | 7.0 | 6.2 | 10.6 | 6.4 | 2.4 | 2.8 | 8.2 | 7.6 | 8.6 | 7.6 |
| Difference/P.E. of difference.. | | | 14.9 | 13.6 | .47 | .02 | 15.0 | 11.4 | 12.07 | 10.92 | 7.3 | 6.7 |
| Difference/P.E. of shorter interval..... | | | 2.18 | 1.79 | .15 | .03 | .88 | .96 | 2.07 | 1.77 | 1.32 | 1.09 |

[1] 104 equals the difference between the means of the first halves of the reaction times obtained at the ½ and 1-second periods; 84 is the difference between the means of the second halves of the same distributions, etc. All probable errors are computed according to the long formula.

The eighth line of this table indicates the size of the differences. It will be seen that the effect of practice is considerable. This is shown by the decrease in the means of the second half as compared with those of the first half of the distribution at each interval. The differences stated in terms of a standard unit, *i.e.*, the probable error of the shorter interval, indicate that there is no constant tendency for the

differences between the means of second to be greater or less than those between the means of the first halves of the series as the interval between the stimuli is increased. However the practice effects is greatest at those intervals which are least favorable for response.

This indicates that the differences found may be due, in part at least, to lack of practice in similar processes at the shorter and longer intervals, or that practice tends to obliterate the refractory phase which follows response.

These results indicate that immediately after responding to a simple auditory stimulus by pressing a telegraph key there is a period of intrinsic unreadiness for response as shown by lengthened reaction times to stimuli applied during this interval. This period is comparable to the refractory phase of more elementary systems.

This period, during which the response to the second of two stimuli is somewhat delayed, is succeeded by a return of the response time to normal, and this in turn by a period of increased readiness for response. This seems to be comparable to the supernormal phase which follows the period of decreased excitability as reported by Adrian and Lucas (1912) and also by Dodge (1927). This period of increased excitability is followed by a subsequent return to approximate normality as shown in Fig. 1. The exact time at which the system returns to normal excitability cannot be determined from the present data. It may be either a rather sudden drop after the 2-seconds interval or a gradual return as shown in Fig. 1. If the return follows a course similar to that of elementary systems, the latter is the more probable. Considerable practice effect is found and tends to reduce the size of the differences between the means at the different time intervals.

*Accuracy in judging linear magnitudes as influenced by the length of the interval between judgments.*—As far as the writer is aware, no one has studied specifically the effect of the length of the interval between successive stimuli on the accuracy of the response. The only studies relating to this problem are those on the general question of the relation of

speed to accuracy. In most of these studies either the speed of making the response or the duration of the stimulus has been varied along with the interval between successive stimuli. This brings in an additional factor which precludes any determination of the influence of the length of the interval alone.

Garrett (1922) gives a rather complete history of the problem of the general relation of speed to accuracy. Interested readers are referred to this study. He also reports some results which bear directly on our problem. He found that in lifting weights and in judging linear magnitudes and handwriting specimens, accuracy tends to increase gradually from the slowest rate used, 4-seconds, to a 2-seconds rate. After this it falls off regularly. The loss becomes more rapid as the speed is increased. This indicates that there is an optimal interval which is close to two seconds for the three processes, and it suggests that there may be an interval at which the factors involved in judgment are at a point of maximal clearness.

In the lifting of weights the time of lifting was always equal to the interval between weight pairs. Garrett made rather limited investigation to determine whether the differences in accuracy at the different rates were due to the varying interval between pairs or to the different lifting rates. He concludes that the evidence indicates that the time of lift is a more important influence in discrimination than the interval.

Certain facts mentioned by this author, however, seem to indicate that the time of lift in judging weights and the length of exposure in judging lines and handwriting specimens may be essentially intervals between judgments. The author noted and the subjects' introspections verified the fact that the decisions were made as soon as the second weight was lifted or the lines and handwriting specimens were exposed. In the case of the latter two the observer's eyes wandered about the room, coming back with a jerk when the bell sounded for the appearance of another card. Garrett mentions that apparently the interval is not a judgment interval but a necessary interim between judgments. If this is so

the real difference throughout may have been one of length of interval between judgments and his conclusion that the differences in accuracy found are not due to the interval is not justified. The similarity between these results and those of the present experiment, in which the interval alone was varied over a considerable range, suggests that this is probably the case.

Our method consisted essentially in presenting pairs of lines for a given period but with different intervals between pairs. The subjects passed judgment on the length of a given one of each pair as to whether it was longer or shorter than the other.

Horizontal lines were ruled at certain prescribed distances on a strip of regular kymograph paper four inches in width. A line was drawn down the center of the paper parallel to its sides, thus dividing each horizontal line into two parts. This gave a series of horizontal lines each of which was divided into two parts by a vertical line running the entire length of the paper. The length of the horizontal lines to the right of the vertical line was always 30 mm. The portions to the left were 33, 31, 29 and 27 mm in length. The lines were drawn on the paper at intervals of $\frac{1}{4}$, $\frac{1}{2}$, 1 inch and 2 inches. Each length of line occurred 20 times, five of which appeared at each of the four intervals used. Thus we had 80 pairs of lines of four different lengths occurring at four given intervals. The lines and intervals were so arranged that each line was presented the same number of times at each interval. They occurred in a random order.

The two ends of the paper were pasted together making it into an endless belt, which was run through a Renshaw polygraph in the usual manner. The speed at which the paper ran was controlled by a Leeds-Worthup constant-speed motor. The rate was so regulated as to move the paper a distance of $\frac{1}{2}$ inch per second. The subject viewed the lines through a slit in a card-board. The opening was $\frac{1}{4}$ by 4 inches in size. Care was taken to insure uniform lighting, and to avoid the casting of shadows across the opening.

The subjects were 25 members of a class in general

psychology at Peabody College, 1 man and 24 women. None of these persons served as subject in the experiment on reaction time.

Each subject was comfortably seated before the opening in which the lines were to appear. He was shown the lines and the opening and given the following instructions: "When the motor is started these lines will appear in the opening before you. As each pair of lines appears you are to judge whether the portion of the horizontal line to the left of the vertical line is longer or shorter than that portion to the right. It it is longer call out long, if shorter call short. If they appear to be the same length, you are to guess. There is always a difference. Be sure that you respond to each pair of lines. You will be given a warning 5 seconds before the first pair appears."

An assistant operated the machine. The writer recorded the responses on a specially prepared blank as the subject called out his judgments. The entire series of eighty pairs was run through four times. Thus each subject made 320 judgments. A 30-seconds rest period was given after each of the four series and a 5-seconds warning was given before the appearance of the first lines of each series.

The results were scored and tabulated according to the number of wrong judgments made on each line at each interval. These results are summarized in Table III.

TABLE III

NUMBER OF ERRORS MADE IN JUDGING LENGTH OF LINES PRESENTED AT DIFFERENT INTERVALS

| Interval in seconds | ½ | 1 | 2 | 4 |
|---|---|---|---|---|
| Percent of correct responses | 93.4 | 96.7 | 95.5 | 95.2 |
| Mean number of errors in 320 responses | 20.94 | 10.58 | 14.36 | 15.40 |
| Standard deviations of errors | 4.00 | 4.42 | 5.11 | 6.72 |
| Difference between successive means | — | 10.36 | 3.78 | 1.04 |
| Probable error of differences[1] | | .63 | .65 | .94 |
| Difference/P.E. of diff. | | 16.44 | 5.82 | 1.11 |
| Diff./P.E. of shorter interval | | 3.84 | 1.27 | .40 |

[1] Probable errors are figured by the long formula.

Since the four lengths of lines appeared the same number of times at each of the four time intervals, comparisons can

be made in terms of the average number of erroneous judgments made at each of the intervals. In plotting the curves found in Fig. 2, it was found that all of the lengths of lines were affected somewhat alike by the changes in time intervals with the exception of the 29 mm lines. The curve for this length showed a very sudden decrease in accuracy at the 2-seconds interval. Examination of the original lines showed that an error had been made in drawing two of these lines. Instead of the variable portions of these lines being 29 mm they were actually 30. This, of course, meant that these lines were judged longer approximately 50 percent of the time. The original data were consequently reworked, eliminating these two lines from the total. Means and standard deviations were multiplied by 10/9 to make them comparable to the others. This gives the results as found under 2 in Table III. The curves in Fig. 2 are constructed from the data given in this table. Omitting the data from these two lines does not influence the general trend of the results, but does bring line 29 into conformity with the general tendency.

Table III shows that accuracy is not a constant function of the length of the interval between successive presentations. It is poorest at the $\frac{1}{2}$-second interval. Where the time between judgments is increased to a 1-second interval there is a decided increase in accuracy, while a further increase to 2 seconds results in a considerable decrease in accuracy. The accuracy at the 4-seconds interval shows a slight further decrease but this is not significant. The difference between the accuracy at the $\frac{1}{2}$- and the 1-second interval is statistically reliable, the critical ratio being 16·44. The difference between the 1-second and the 2-seconds intervals is also significant, being 5.82 times its probable error. The difference between the mean number of errors at the 2- and 4-seconds intervals is negligible. The last row of Table III gives the differences in terms of the probable error of the shorter of each of the two intervals compared. A comparison of the $\frac{1}{2}$- and 2-seconds intervals (not in table) shows a difference of 2.44 probable error units in favor of the latter. The corresponding measures for the $\frac{1}{2}$- and 4-seconds intervals

TABLE IV

MEAN NUMBER OF ERRORS AND PERCENT ACCURATE ON EACH LENGTH OF LINE AT THE FOUR TIME INTERVALS

| Interval | ½ second | | | | 1 second | | | | 2 seconds | | | | 4 seconds | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Length of line[1] | 27 | 29 | 31 | 33 | 27 | 29 | 31 | 33 | 27 | 29 | 31 | 33 | 27 | 29 | 31 | 33 |
| Mean number of errors | 3.44 | 9.80 | 5.20 | 2.00 | 1.28 | 3.64 | 4.04 | 1.12 | 1.20 | 5.20 | 5.36 | 2.20 | 1.88 | 4.48 | 5.96 | 2.84 |
| Percent accurate | 95.7 | 87.8 | 93.5 | 97.5 | 98.4 | 95.4 | 95.0 | 98.5 | 98.5 | 93.5 | 93.3 | 97.3 | 97.7 | 94.4 | 92.6 | 96.5 |

[1] Standard for comparison was always 30 millimeters.

gives a difference of 2.05 probable error units in favor of the longer interval.

The average number of errors made on each length of line at each of the four intervals is given in Table IV. The percentage accuracy is also included. In order to see how far the judgments on the different lengths follow the general tendency of the entire group, the curves in Fig. 2 were drawn as previously mentioned. It will be noticed that they all

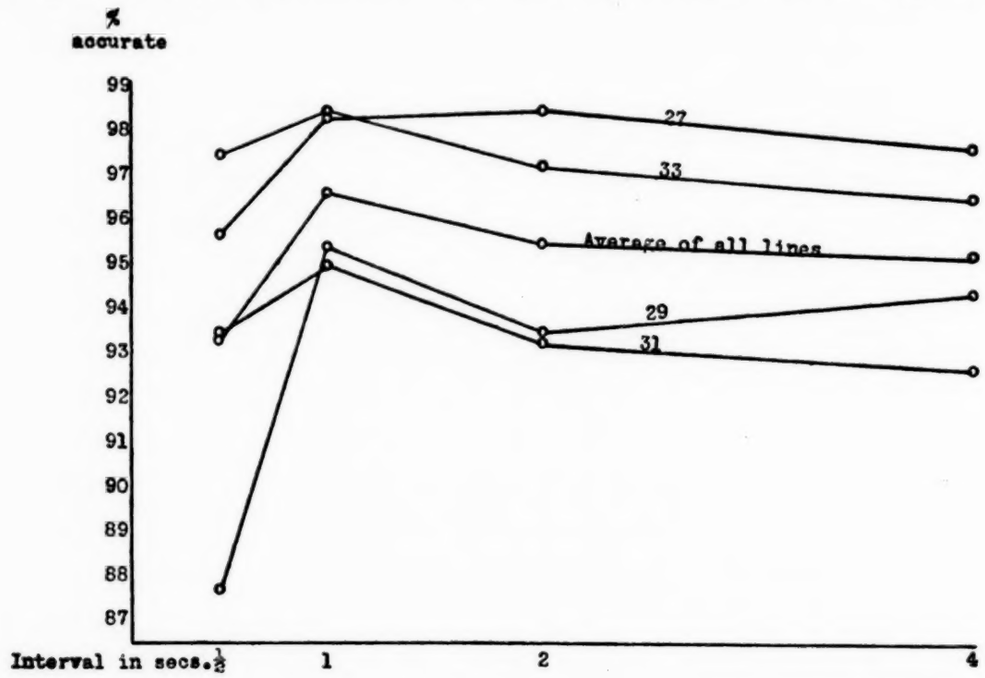



FIG. 2. Accuracy curves for judging lengths of lines with varying time-intervals between judgments

have the same general form and conform quite well to the general tendency as shown by the composite curve for all lines, which is given for comparison.

There were totals of 195, 578, 514 and 204 errors made in estimating the 27, 29, 31, and 33 millimeter lines respectively. This gives a total of 773 errors on those lines which were shorter than the standard and of 714 on those longer than the standard, and it indicates a slightly greater accuracy for the longer lines. These results agree with those reported by Garrett (1922). It will be noted that there was a high degree of accuracy at all intervals for the four lengths used. It would probably have been better if the differences had

approached more nearly the subjects' limen so as to give nearer 75 percent correct responses. Great individual differences, of course, make it difficult to produce this condition.

There is a striking similarity between the results reported here and those found with reaction times. In judging linear magnitudes there is an optimal interval at which accuracy is greatest. Varying this interval in either direction results in a considerable increase in errors. Reducing the inter-period exposures from 1 to $\frac{1}{2}$ second doubles the number of errors in judgments. Increasing the interval from 1 to 2 seconds results in a 32.89 percent increase in errors. Increasing the interval to 4 seconds increases the number of errors to 45.56 percent over the number at 1 second.

Again we find evidence of an intrinsic unreadiness for response immediately after responding. This is indicated by the large percentage of errors at the $\frac{1}{2}$-second interval. This is followed within the next $\frac{1}{2}$-second by a period during which accuracy is considerably above that of longer periods. Such results seem to suggest the operation of a period comparable to that found in simpler mechanisms following the refactory period.

Our results agree with those of Garrett (1922) on the relation of speed to accuracy.

In the course of this experiment each subject gave judgments on the same series of eighty lines four times. To determine fatigue and practice effects, the mean number of errors made and the standard deviations for the 25 subjects on each of the four series was computed. These results are shown in Table V.

As a measure of these effects the mean number of errors made on the first series minus the mean number on the fourth series divided by the mean of the first has been used. This is shown in the fifth line of the table. The sixth line gives the difference in terms of the standard deviation of the first series. There seems to be some evidence of a practice or fatigue effect which operates somewhat differentially at the several intervals. At the two shorter intervals more errors are made on the fourth then on the first trial while

TABLE V

COMPARISONS OF THE PRACTICE AND FATIGUE EFFECTS FOUND IN JUDGING LENGTHS OF LINES WITH VARYING INTERVALS BETWEEN JUDGMENTS

| Interval in secs. | ½ | | | | 1 | | | | 2 | | | | 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Series | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Mean no. of errors | 5.54 | 5.94 | 5.14 | 6.02 | 2.66 | 2.94 | 3.02 | 3.36 | 3.90 | 3.38 | 3.30 | 3.70 | 4.94 | 3.70 | 4.38 | 4.18 |
| S.D. | 2.44 | 1.96 | 2.05 | 1.65 | 1.44 | 1.70 | 1.47 | 1.65 | 1.89 | 1.54 | 1.36 | 1.74 | 2.08 | 2.21 | 2.12 | 2.28 |
| $\frac{\text{F.} - \text{L.}}{\text{F.}}$ | −.09 | | | | −.27 | | | | .05 | | | | .15 | | | |
| $\frac{\text{F.} - \text{L. of F.}}{\text{S.D.}}$ | −.20 | | | | −.50 | | | | .11 | | | | .37 | | | |

with the two longer intervals the opposite is true. The increase in errors at the shorter intervals is what one would expect from fatigue accompanied by an increased refractory period. This increase may not be sufficiently great to make itself felt at the longer intervals. The decrease in errors at the longer intervals may indicate that a period of hyper-excitability is being shifted in that direction.

*Repetitive tendencies in an associative process as influenced by the length of interval between responses.* Thorndike (1927) reported results which seem to indicate something similar to a refractory period in associative processes. The experiment was designed primarily to throw light on the influence of repetition. In the experiment six subjects wrote a number (from 0 to 9) whenever they heard a word read by the experimenter. By this method 3840 words were read at the rate of one every $2\frac{1}{2}$ seconds. The subjects were told that when the experimenter read a word they would think of some number from 0 to 9, and that they should write the first of these numbers which came to their minds. If none of the subjects had favored numbers and chance alone determined which numbers were written, there should be ten percent of sequences like 00, 11, 99, 55, etc. The actual number of sequences is far below this number.

In a second experiment of the same general nature, but with a time interval of 5 seconds between successive readings of the words, the avoidance of sequences was not nearly so great. An analysis of Thorndike's data shows that in 15,360 responses to words read at the rate of one every $2\frac{1}{2}$-seconds, there were 663 or 4.32 percent of sequences. Of 5120 responses at the 5-seconds intervals there were 421 or 8.22 percent of sequences. Doubling the interval almost doubled the percent of sequences.

Thorndike suggests that the increased avoidance of sequences as the interval is shortened may be due, in part at least, to a refractory period. Dodge (1927*a*) in a note on Thorndike's experiment seems to consider this interpretation probable. He also suggests a crucial extension of Thorndike's experiment. This is simply to use various time intervals

between stimuli. An actual refractory phase as a barrier against repetition should theoretically be more apparent the nearer the responses approach each other in time. Furthermore, if the tendency noted by Thorndike is really due to a refractory phase one would expect, other things being equal, to find a critical frequency at which the tendency to repetition would be equal to expectation by chance. Then would follow a subsequent period of rebound during which the tendency to repeat would be exaggerated and a return to approximate normality.

The present experiment is essentially a repetition and extension of Thorndike's experiment, as suggested by Dodge.

*Method.*—Our method was practically the same as Thorndike's. Nonsense syllables, consisting of a vowel between two consonants, were read at varying rates. Numbers from 1 to 10 inclusive were written as associative responses, which were analyzed as to the number and percentage of sequences occurring with the various time intervals.

The time intervals were regulated by a muffled metronome. Responses were recorded by the subjects on specially prepared blanks.

The subjects were each given a blank upon which the responses were to be recorded and then given the following instructions: "I am going to read a list of nonsense syllables. Every time I read a syllable you will think of a number between 1 and 10, inclusive. Each time I read a syllable you are to write the number you think of in the spaces provided for that purpose. Always put down the first number that comes to your mind. I shall read the syllables at varying speeds, some rapidly and some slowly. First, I shall read one every two seconds." They were then instructed as to the method of recording their responses, the metronome was started and the syllables were read. With three girls in a preliminary experiment, 1-, $1\frac{1}{2}$-, $2\frac{1}{2}$-, 5-, $7\frac{1}{2}$-, and 10-seconds intervals were used. The adults in the main experiment responded at 1-, 2-, 4-, 8-, and 16-seconds intervals. Increasing the periods beyond these bored the subjects so that experimentation was difficult. In order to get as nearly

as possible a double fatigue order, the following arrangement of intervals was used with the adults: 2 seconds, 8 seconds, 16 seconds, 1 second, and 4 seconds. A similar order was followed with the three girls.

Preliminary work was carried out with three girls, aged 8, 10, and 11 years respectively. The work extended over a period of three weeks at irregular intervals. These results are given in Table VI. In this and succeeding tables,

TABLE VI

REPETITIVE TENDENCIES IN THE ASSOCIATIVE RESPONSES OF 3 GRAMMAR SCHOOL GIRLS

| Interval (sec.) | Subj. | No. responses | No. sequences | Percent of sequences | Number of series | Number of responses minus number series | Percent of sequences with series eliminated |
|---|---|---|---|---|---|---|---|
| 1 | A[1] | 250 | 4 | 1.6 | 56 | 194 | 2.1 |
| | E | 150 | 3 | 2.0 | 23 | 127 | 2.4 |
| | L | 200 | 1 | .5 | 39 | 161 | .6 |
| | Total | 600 | 8 | 1.3 | 118 | 482 | 1.7 |
| 1½ | A | 454 | 8 | 2.0 | 130 | 324 | 2.4 |
| | E | 496 | 8 | 1.6 | 91 | 405 | 2.0 |
| | L | 509 | 20 | 3.9 | 150 | 359 | 5.6 |
| | Total | 1459 | 36 | 2.5 | 371 | 1088 | 3.3 |
| 2½ | A | 478 | 18 | 3.8 | 119 | 359 | 5.0 |
| | E | 408 | 12 | 3.0 | 59 | 349 | 3.2 |
| | L | 484 | 7 | 1.5 | 111 | 373 | 1.9 |
| | Total | 1370 | 37 | 2.7 | 289 | 1081 | 3.4 |
| 5 | A | 614 | 63 | 10.1 | 105 | 509 | 12.5 |
| | E | 592 | 32 | 5.4 | 74 | 518 | 6.2 |
| | L | 598 | 88 | 14.7 | 101 | 497 | 17.7 |
| | Total | 1804 | 183 | 10.1 | 280 | 1524 | 12.0 |
| 7½ | A | 521 | 116 | 22.2 | 72 | 449 | 25.6 |
| | E | 475 | 77 | 16.3 | 61 | 414 | 18.6 |
| | L | 524 | 131 | 24.0 | 75 | 449 | 29.1 |
| | Total | 1520 | 324 | 21.2 | 208 | 1312 | 24.8 |
| 10 | A | 351 | 90 | 25.7 | 33 | 318 | 28.2 |
| | E | 350 | 99 | 28.2 | 35 | 315 | 31.2 |
| | L | 400 | 125 | 31.2 | 47 | 353 | 35.4 |
| | Total | 1101 | 314 | 28.4 | 115 | 986 | 32.0 |

[1] A, E and L refer to the three girls. A is age 8, E age 10 and L age 11 years.

'sequences' refers to numbers succeeding themselves as responses, for example, 44, 11, 77, etc. When the same number succeeded itself twice as a response as 333, 555, etc., it was counted as two sequences, four in a row, as three sequences, etc. Numbers in 'series' are those occurring in the same order as they do in the number series as, 123, 567, etc. In such a case, 1 and 5 would not be in series while 23 and 67 would.

The main part of the present experiment was carried out with 31 members of a graduate course in educational psychology at George Peabody College for Teachers (11 men and 20 women). Eleven of these persons had previously served as subjects in the reaction time experiment.

*Results.*—In tabulating the results it was found that as the time interval between stimuli decreased there was a decided increase in the tendency to count. It was thought that any tendency to avoid sequences at the shorter intervals might really be a greater tendency to fall into the habit of counting. Therefore it became necessary to determine the extent of the counting and see if the avoidance of sequences was still found at the shorter intervals with the numbers in series eliminated. These results are given and comparisons made in Tables VI and VII.

TABLE VII

REPETITITIVE TENDENCIES IN THE ASSOCIATIVE RESPONSES OF 31 GRADUATE STUDENTS

| Interval in seconds | 1 | 2 | 4 | 8 | 16 |
|---|---|---|---|---|---|
| Number of responses | 1860 | 1891 | 1550 | 1488 | 1240 |
| Number of sequences | 58 | 117 | 111 | 93 | 121 |
| Percent of sequences | 3.12 | 6.19 | 7.16 | 6.25 | 9.76 |
| Number of responses in series | 304 | 352 | 234 | 220 | 206 |
| Number of responses minus series | 1156 | 1549 | 1316 | 1268 | 1034 |
| Percent of sequences, series eliminated | 3.72 | 7.55 | 8.43 | 7.33 | 11.70 |
| Mean number of sequences per 50 responses | 1.86 | 3.77 | 4.22 | 3.66 | 5.85 |
| Standard deviations of sequences | 2.51 | 4.02 | 4.93 | 3.54 | 4.41 |
| Difference between successive means | — | 1.91 | .45 | .56 | 2.19 |
| Probable error of difference [1] | — | .45 | .55 | .46 | .37 |
| Critical ratio. Diff./P.E. diff. | — | 4.24 | .82 | 1.22 | 5.92 |
| Diff./P.E. of shorter interval (size of diff. in terms of standard units) | — | 1.13 | .17 | .17 | .74 |

[1] Long formula was used in getting probable errors of differences

$r_{12} = .42$; $r_{24} = .24$; $r_{48} = .64$; $r_{8\,16} = .72$.

An examination of these tables shows a general tendency for the percentage of sequences to increase as the interval between responses is lengthened. This increase is, however, not proportional. The three girls show a lower percentage of sequences at the shortest interval used and a much more rapid increase as the interval is increased. To facilitate comparisons the results are shown graphically in Fig. 3.

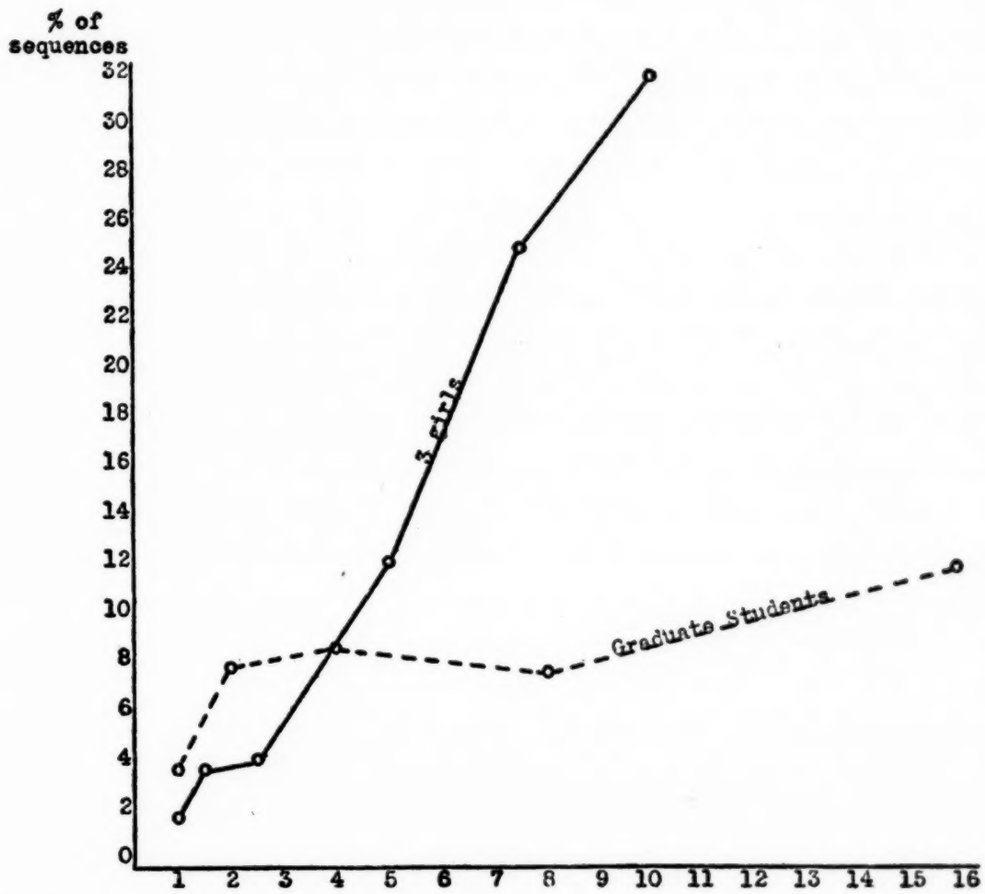



FIG. 3. Percent of sequences with different time intervals between stimuli.

Here the difference between the rates of increase for the three girls and for the graduate students is shown very markedly. The girls have an increase from 1.7 percent of sequences at the one second interval to 32.0 percent at an interval of ten seconds. The 31 graduate students show an increase from 3.72 at the 1-second to 11.70 percent at the 16-seconds interval.

The results of this section differ somewhat from those of

the two preceding. The principal difference is that, within the limits of the time intervals used, there is a general tendency for the perseverative responses to increase as the interval between stimuli is lengthened, with no evidence of a subsequent drop. The period over which the increase extends is also very much prolonged beyond that found for reaction times and judgments. It is very probable that if the time intervals were made sufficiently long a period will be found beyond which there is no further increase or a decrease in the tendency to repeat. The results show that the percentage of sequences is much greater with the three girls than is the case with the graduate students. The number in the one group is, of course, too small to warrant any conclusions as to differences in general. The consistency of the results, however, suggests that such a difference may exist.

Something very similar to a refractory period seems to be operative here. It is more evident in the case of the three girls than in that of the graduate students. Immediately after responding by writing one number there is a decided tendency against responding again in the same way. Not only is there a tendency for numbers not to occur in sequences, but there is also a marked tendency to use a greater variety of numbers in response in the case of short intervals. These are not the same thing. For example, one could avoid sequences while limiting his range of responses to only two numbers. Figuring the number and percentage of 1, 2, 3, etc., occurring at each interval shows that a greater range of numbers is used at the shorter intervals. As the intervals increase there is a greater tendency to limit response to a few favored numbers. It is impossible to determine from the present data the extent of the period during which there is a refractoriness of the system involved. Thorndike (1927) suggested that when nothing is operating for or against repetition we should find 10 percent of sequences. We shall actually find more than this because of the subjects' having favored numbers, and certain habits. The ten percent standard would give about 5 seconds for the girls and somewhere between 8 and 16 seconds for the graduate students as

the intervals which give the number of sequences expected by chance. Such figures have but little value, however.

Thorndike (1927) found that the tendency to avoid sequences increased with practice. Our series in the case of the adults were far too short to show any practice effects. Since each of the three girls in the preliminary experiment gave several hundred responses at each interval their results may show some practice effects. The average number of sequences written by the three girls at each time interval during the first one hundred and the last one hundred responses are given in Table VIII.

TABLE VIII

MEAN NUMBER OF SEQUENCES MADE BY THREE GIRLS ON THE FIRST AND LAST 100 RESPONSES AT EACH TIME INTERVAL

| Interval in seconds | 1 | 1½ | 2½ | 5 | 7½ | 10 |
|---|---|---|---|---|---|---|
| Mean number of sequences in first 100 responses | .67 | 4.25 | 2.67 | 11.67 | 18.33 | 22.66 |
| Mean number of sequences in last 100 responses | 1.33 | 1.00 | 3.33 | 12.00 | 20.00 | 23.33 |
| First minus last / First | −1.00 | .77 | −.25 | −.03 | −.09 | .01 |

This table shows no consistent change with practice. Two intervals show decreases in the number of sequences and four show increases. These results do not agree with Thorndike's. The small number of cases and the difference in age of the two groups make comparisons with Thorndike's data of little value.

## III. INTERPRETATIONS AND CONCLUSIONS

The evidence which we have presented shows that simple reaction time to an auditory stimulus is considerably lengthened when the stimulus follows a preceding one at less than a certain interval of time. This interval is between ½ and 1 second in length. Moreover, accuracy in judging the length of lines is found to be considerably decreased if the response occurs before a certain interval has elapsed since a previous judgment. The period during which there is a return to normal seems also to be between ½ and 1 second in length

in this process. Finally, there is a decided tendency against immediate repetition in simple associative processes. This tendency, according to our data, decreases as the interval between responses increases up to a 10-seconds interval in the preliminary case of three grammar school girls, and it decreases up to a 16-seconds interval in that of 31 graduate students, our main experimental group. These are the longest intervals used in the present experiment. These three activities, then, all show evidence of a tendency against immediate repetition.

Accuracy in judging length of lines and the speed of simple reactions are greatest when there is approximately a 1-second interval between responses. When the period between responses is increased to 2 and 4 seconds there is a progressive decrease in accuracy of judgment and an increase in reaction time.

We shall now take up the possible explanations of these results. There is a striking similarity between the results of our reaction-time experiments and those obtained by Woodrow (1914). He was interested primarily in the effect of the length of the preparatory interval on reaction time. His results show that a regular interval of about 2 seconds between the giving of the preparatory signal and the stimulus for response is most favorable for a quick response. As the interval is shortened below this favorable region, or lengthened beyond it, the reaction time increases. This increase continues up to intervals of over 24 seconds. When the length of the preparatory interval was varied irregularly, he found that the most favorable period of warning for a quick response was about 12 seconds. Varying the interval either way from 12 seconds increased the average reaction time. When two subjects gave serial reactions to each of a series of regularly recurring sounds, the most favorable interval was 3.5 seconds for one and 2 seconds, the shortest period used, for the other subject. These latter results, of course, differ from ours in that Woodrow's intervals were regular while ours were irregular.

The regular intervals were tried in certain preliminary

experiments in our own study. We found that at the shorter intervals, responding to a regularly recurring sound resolved itself for the reactors to simply a process of 'beating time.' The response soon came to be given prematurely or at the same time as the stimulus. There was no reaction time as such really measured. It is evident that in such reactions to regular periodic stimuli the situation is quite different from one in which the subject does not know when the stimulus is to be given and must be ready at all times. When the serial responses follow each other rapidly and at regular intervals each succeeding response comes gradually to be determined by the proprioceptive impulses set up by the previous response. In such a case the response may occasionally be initiated even before the auditory stimulus is received. That this is the case seems evident from the fact, observed by us, that if in any single case the stimulus interval is prolonged unexpectedly, the subject responds at approximately the same interval as formerly. The absence of the exteroceptive stimulus has under such circumstances little, if any, effect. We did not find a subsequent increase in reaction time as the interval between responses was increased beyond 4 seconds. Our data are inadequate on this point, however, and cannot be taken as conclusive.

Woodrow (1914) considered the prolongation in reaction time by unfavorable (*i.e.* very short or very long) intervals to be due to attention. He recognized, however, that attention is a process of neuro-muscular adjustments or accommodations. To call it 'attention' does not explain it. This is but a process which must itself be explained in terms of neuro-muscular sets and tensions. A refractory period and supernormal phase, such as we have suggested, would probably play a very important part in the release of tensions and recovery of sets.

In such complex systems as reflexes, the period of diminished excitability following response may be several seconds in duration. This is much longer than can be accounted for on the basis of the recovery of the elementary nerves and muscles involved in the response when considered in isolation.

Most reflex and voluntary responses consist of at least a volley of impulses in the neurones and a tetanic response in the muscles (Forbes, Querido, Whittaker and Hurxthal, 1928; Travis, Tuttle, and Hunter, 1927). The action currents set up in the case of voluntary movements are essentially the same as those in reflex responses (Travis, Tuttle, and Hunter, ibid.). The prolonged refractory period in such systems must be an interval long enough for the return of the equilibrium of the system as a whole. It is probably a period during which there is a reinstatement of the particular sets and tensions, the release of which makes the particularre sponse possible and most efficient.

Considered in this light, the refactory phase may be a very important factor in those sets and changes in sets known as 'attention' and 'fluctuations of attention.' This means that explanations such as the one Woodrow has offered are not opposed to one stated in terms of the post-stimulation phenomena of more elementary systems.

The inertia of the reacting organ could hardly have played any significant part, since our own data as well as those of Dresslar (1892) and Dodge (1927) show that such responses can be made at the rate of from 6 to 12 per second.

Some such 'attentional' factor might also be evoked to explain the decrease in accuracy of judging lines when the interval was varied in either direction from about 1 second. Unless 'attention' is a spontaneous faculty, however, its phenomena must be explained in such a way as is suggested above.

There is, however, one possible source of error in the judging of the lengths of lines. This is in the matter of articulation at the $\frac{1}{2}$-second interval. The subjects reported —and the experimenter's observation substantiated the report —that they had difficulty in saying the right word at the shorter intervals. Sometimes they reported that they wanted to say one thing but said another. Garrett (1922) observed the same tendency in his subjects, who gave six different kinds of responses. He attributed this difficulty to the fact that all six responses were equally ready to be given and the

slightest cue often set off a particular one which closer observation proved to be wrong. Our results show that limiting the number of possible responses to two, longer or shorter, does not remove the difficulty.

One of our subjects responded by touching one key with his right hand when the line was judged longer and another key with his left hand when it was judged shorter. Under these circumstances he made almost double the number of errors as compared with his record made when calling aloud the response. This evidence, though not conclusive, suggests that the difficulty may be more fundamental than mere vocalization. The subject was not accustomed to responding by pressing keys in the prescribed manner and practice would doubtless have reduced his errors somewhat. If difficulties in vocalization did bring about the increased errors at the shorter intervals, an additional factor must be evoked to account for the greatest accuracy at a period slightly greater than this and the subsequent decrease in accuracy as the interval is prolonged.

There is one other difficulty in judging the length of lines which must be considered. This is the matter of the order in which the lengths appeared. No other worker seems to have considered this factor. It is very probable that the judgment of a given pair is a function of the length of the preceding lines as well as of the two which are presented simultaneously at any time. With lengths such as we have used, a line 31 mm long would probably be judged correctly more often if it followed a 27 mm than if it followed a 33 mm line. Such a holdover effect probably varying roughly in inverse relation to the length of the period between exposures would explain the increase in errors at the $\frac{1}{2}$-second interval. The fact that accuracy is greatest not at the longest interval, as would be expected from such an hypothesis, but at one which is only slightly longer than the one where accuracy is poorest indicates that some additional factor is operative. The order of the appearance of the lines was determined by chance. The arrangement used here is the one more commonly met with in ordinary behavior of this sort.

For purposes of explanation however, it makes considerable difference whether the inaccuracy at the ½-second interval is due to the order of appearance of the lines or to the fact that the reacting system is refractory due to the previous response. In as much as it is a holdover effect from the previous exposure, it is a post-stimulation phenomenon but it is hardly an explanation in terms of a refractory phase as this is ordinarily understood.

Dodge (1927*a*) suggests three plausible hypotheses to account for the avoidance of sequences in associative responses. First, the subjects may have developed, by some unknown condition, a temporary bias in favor of the distribution of their responses. This might have arisen as a chance interpretation of the instructions. Secondly, the general tendency might represent a habit common to all of the members of the group. The third possible explanation is in terms of a prolonged refractory phase. After considering these possibilities, Dodge concludes that the unanimity and consistency of Thorndike's results suggest some underlying tendency such as a refractory period to account for it.

In the present experiment the instructions were worded so as to emphasize the fact that the response was to be purely an associative one; that they were not to be selected. The first number coming to mind was to be written. In spite of this, several of the graduate students reported that they were aware of a tendency to vary their responses. Some of them also reported that they soon came to associate certain vowel sounds with particular numbers, for example, a with 1, e with 2, etc.

None of the three girls reported any such tendencies in our preliminary work. The girls did notice that they were writing the same number a great many times at the longer intervals. They spontaneously remarked about it several times. One of them said to another once during the experiment, "I know he is going to get after me for writing so many sevens." If such a bias does explain the small number of sequences, it did not effect all intervals proportionally. If it affected the graduate students more than the children it

offers a possible explanation of the differences in percent of sequences at the longer intervals. It is obvious that any permanent associations, such as certain graduate students reported, would tend to decrease any effect the change in length of interval might have. This may also explain the small increase in percentage of sequences with increase in interval in the results of the larger group as compared with those of the children.

We have now examined several possible explanations of our experimental results. None of them has been found a complete explanation of our data. It seems probable that the effect of a refractory period, followed by a period of hyper-excitability and a subsequent return to normal in the systems involved, is worth considering seriously as a large factor in the explanation of these phenomena.

We do not wish to imply that we consider this to be an ultimate explanation. The refractory and hyper-excitable periods are but names given to the phenomena which must themselves be explained. It is only one step forward in our understanding of the many factors conditioning human behavior.

*Practical applications.*—The question of the relation of the accuracy of a judgment or the speed of a response to the time intervening since a previous response was made is of no little importance. No one would question the desirability of accuracy. On a purely *a priori* basis one would hardly expect to find any constant relationship between the accuracy of a judgment and the interval between this and a previous response, unless such an interval were extremely short. The relationship between accuracy and speed is generally considered to be an inverse one.

Our results indicate that in certain processes, at least, there is an optimal interval for the accuracy of successive judgments of certain kinds. Varying this interval either way from the optimal value results in a loss of accuracy. It is probable that the length of this interval will vary both with the individual and with the nature of the response. The same thing seems to be true of the speed of reaction.

This suggests that in industry where rapid movements or accurate judgments are required at certain intervals, the length of the interval may have considerable influence upon the speed or accuracy of the resulting response. The optimal interval for any particular response could easily be determined experimentally.

Our findings on the relation of the tendency-to-repeat to the length-of-interval-between-responses may have considerable significance for the problem of learning. Our results are in agreement with those of Peterson (1917), Kuo (1922), Lashley (1924), Thorndike (1927) and Dunlap (1928) on the influence of 'recency' in learning. Within the limits of the intervals here used, there is an inverse relationship between recency of response and the liklihood of the same response occurring again. This is the exact reverse of the 'law of recency' as it is usually stated. Our results show that in the case of the 3 girls the chances are on the average 24.6 to 1 greater that a given response will follow itself if the second response occurs 10 seconds after a first than if only 1 second has elapsed. In the case of the 31 graduate students the similar chances of a repetition are 3.12 times greater at the end of a 16-second interval than of a 1-second interval. This may have some relation to the superiority of spaced over unspaced practices in learning, and it may also throw some light on the troublesome problem of the elimination of 'errors' in learning.

In answer to the original questions of the present study we can now say that: (1) Voluntary responses, judgments and simple associative responses do produce effects in the organism which serve as a barrier against immediate repetition. (2) The subsequent recovery and return to approximate normality, as it manifests itself in the ways here investigated, is comparable to that found in isolated tissues and reflexes. It may not be identical with the latter but based upon it.

*Suggestions for further study.*—There are several questions growing out of the present study which could well be investigated.

1. The relation between accuracy and the length of the interval in responses of various kinds.

2. An extension of the present study of repetitions in associative responses using (1) greater time intervals, (2) different kinds of material and forms of response, (3) subjects of different ages and levels of ability. The writer intends to extend the work along these particular lines.

3. Detailed study of the relation of the distribution of practice to rate of learning with particular reference to the course of recovery of excitability in more elementary systems.

BIBLIOGRAPHY

ADRIAN, E. D., AND LUCAS, K. (1912). On the Summation of Propagated Disturbances in Nerve and Muscle, *J. Physiol.*, **44**, 68–124.

BAZETT, H. C. (1908). Observation on the Refractory Period of the Sartorius of the Frog, *J. Physiol.*, **29**, 231–340.

BOYCOTT, A. E. (1899). Note on the Muscular Response of the Sciatic Nerve to Two Stimuli, *J. Physiol.*, **24**, 144–154.

BROCA, A., ET RICHET, C. (1897). Period réfractaire dans les centres nerveux, *C. R. Acad. Sci.*, **124**, 96–99.

COHEN, L. H. (1929). The Relationship between Refractory Phase and Negative Adaptation in Reflex Responses, *J. Comp. Psychol.*, **9**, 1–16.

DENNY-BROWN, D. E., AND LIDDELL, E. E. T. (1927). Observations on the Motor Twitch and on Reflex Inhibition of the Tendon-Jerk of M. Supraspinalis, *J. Physiol.*, **63**, 7–79.

DODGE, R. (1911). Exploration of a Normal Knee-jerk, *Zsch. f. All. Physiol.*, **12**, 1–55.

DODGE, R. (1913). Refractory Phase of the Normal Wink Reflex, *Am. J. Psychol.*, **24**, 1–7.

DODGE, R. (1926). The Problem of Inhibition, *Psychol. Rev.*, **33**, 1–33.

DODGE, R., AND LOUTTIT, C. M. (1926*a*). Modification of the Pattern of the Guinea Pig's Reflex Response to Noise, *J. Comp. Psychol.*, **6**, 267–285.

DODGE, R. (1926*b*). Theories of Inhibition, Part II, *Psychol. Rev.*, **33**, 168–187.

DODGE, R. (1927). *Elementary Conditions of Human Variability*, Columbia University Press.

DODGE, R. (1927*a*). A Note on Professor Thorndike's Experiment, *Psychol. Rev.*, **34**, 237–240.

DRESSLAR, F. B. (1892). Some Influences which Affect the Rapidity of Voluntary Movements, *Am. J. Psychol.*, **4**, 1–14.

DUNLAP, K. (1928). A Revision of the Fundamental Law of Habit Formation, *Science*, **67**, 360–362.

ENGELMANN, H. W. (1895). Boeobachtungen und Versuche am suspendierten Herzen. Dritte Abbandlung. Réfractaire Phase und Compendsatorische Ruhe in irher Bedeutung für Herzrythmus, *Pflüg. Arch. f. d. ges. Physiol.*, **59**, 309–349.

FORBES, A., RAY, L. H., AND GRIFFITH, F. R., JR. (1923). The Nature of the Delay in the Second of Two Stimuli in Nerve and in the Nerve-Muscle Preparation, *Am. J. Physiol.*, **63**, 553–617.

FORBES, A. (1929), in *Foundations of Experimental Psychology*, edited by Carl Murchison, pp. 139, 140, Clark University Press.

FORBES, A., QUERIDO, A., WHITAKER, L. R., AND HURXTHAL, L. M. (1928). Electrical Studies in Mammalian Reflexes. V. The Flexion Reflex in Response to Two Stimuli as Recorded from the Motor Nerve, *Am. J. Physiol.*, **85**, 432–457.

GARRETT, H. E. (1922). A Study of the Relation of Accuracy to Speed, *Arch. Psychol.*, **8** (No. 58).

GASSER, H. S., AND ERLANGER, J. (1925). The Nature of Conduction of an Impulse in the Relative Refractory Period, *Am. J. Physiol.*, **73**, 613–635.

GATEWOOD, E. L. (1920). Individual Differences in Finger Reactions, *Psychol. Monog.*, **28**, (No. 126).

GOTCH, F. (1910). The Delay of the Electrical Response of Nerve to a Second Stimulus, *J. Physiol.*, **40**, 256–274.

GOTCH, F., BURCH, G. J. (1899). The Electrical Response of Nerve to Two Stimuli, *J. Physiol.*, **24**, 410–426.

HERMANN, L., AND UMRATH, K. (1927). Über das Refracturstaduim des Herzens, *Pflüg. Arch. f. d. ges. Physiol.*, **215**, 365–372.

JOHNSON, H. M. (1923). Reaction-Time Measurements, *Psychol. Bull.*, **20**, 562–589.

KUO, Z. Y. (1922). The Nature of Unsuccessful Acts and their Order of Elimination in Animal Learning, *J. Comp. Psychol.*, **2**, 1–27.

LASHLEY, K. S. (1924). Studies of Cerebral Function in Learning. VI. The Theory that Synaptic Resistance is Reduced by the Passage of the Nerve Impulse, *Psychol. Rev.*, **31**, 369–375.

LUCAS, K. (1909). On the Refractory Period of Muscle and Nerve, *J. Physiol.*, **39**, 231–340.

LUCAS, K. (1911). On the Recovery of Muscle and Nerve after the Passage of a Propagated Disturbance, *J. Physiol.*, **41**, 368–408.

LUCAS, K., AND ADRIAN, E. D. (1917). *The Conduction of the Nervous Impulse*, Longmans Green.

MAREY, J. E. (1876). Des excitations Artificielles du Cœur, *Travaux du Laboratorie de M. Marey*, **II**, p. 63.

MCWILLIAM, J. A. (1888). On the Rhythm of the Mammalian Heart, *J. Physiol.*, **9**, 167–198.

PETERSON, J. (1917). Frequency and Recency Factors in Maze Learning by White Rats, *J. Animal Behav.*, **7**, 336–364.

SCHULTZ, W. H. (1906). The Effect of Chloralhydrate upon the Properties of Heart Muscle, *Am. J. Physiol.*, **16**, 483–498.

SHERRINGTON, C. S. (1906). *The Integrative Action of the Nervous System*, Yale University Press.

STRUGHOLD, H. (1927). Beiträge zur Kerntnis des Refraktärphasen des Menschlichen Rückenmarkes, *Zsch. Biol.*, **85**, 453–470.

THORNDIKE, E. L. (1927). The Refractory Period of Associative Processes, *Psychol. Rev.*, **34**, 234–236.

THORNDIKE, E. L. (1927*a*). A Fundamental Theorem in Modifiability, *Proc. Nat. Acad. Sci.* **13**, 15–18.

TRAVIS, L. C., TUTTLE, W. W., AND HUNTER, T. A. (1927). The Tetanic Nature of the Knee-jerk Responses in Man, *Am. J. Physiol.*, **81**, 670–678.

VERWORM, M. (1913). *Irritability*, Yale University Press.

WOODROW, H. (1914). The Measurement of Attention, *Psychol. Monog.*, **17** (No. 76).

ZWAARDEMAKER, H. (1904). Sur Une Phase Réfractaire Du Réflexe de Deglutition, *Arch. Int. de Physiol.*, **1**, 1–16.

ZWAARDEMAKER, H., UND LANS, L. J. (1899). Ueber ein Stadium relativer Unerregbrakeit als Ursache des intermittirenden Charaktere des Lidschlagreflexes, *Centbl. f. Physiol.*, **13**, 325–329.

(Manuscript accepted March 21, 1930)

# THE RELATION OF MEMORY TO INTELLIGENCE [1]

BY EURI BELLE BOLTON, PH.D.

*Georgia State College for Women*

I. *Introduction.*—In his earlier attempts to measure individual differences in intelligence, Binet (1897 and 1911, pp. 162–170) considered memory functions to be most important factors, but later he thought them to be independent of judgment and hence of less importance. He used memory tests as measures of mental ability because they measure the higher thought processes, and in his 1908 scale for the measurement of intelligence, 23, or 39.7 percent, of the 58 tests are tests of some sort of memory function (Peterson, 1925, p. 193 f.). In the Stanford revision of the Binet scale there are 90 tests, including those that are alternatives, and 33 or 36.7 percent of these are tests of recognition or of some memory function; that is, tests which do not require any reorganization of knowledge (Terman, 1916, pp. 142–348). The Binet test is generally regarded as one of the best criteria for evaluating intelligence tests. For this reason it is important to determine the relation between memory on the one hand and intelligence as measured by tests other than memory tests on the other.

## II. *Investigations of the relation of memory to intelligence.*

From the results of a study of the fatigue effects of school children made in Breslau in 1897, Ebbinghaus concluded that memory tests have little value in determining

[1] A fellowship granted by the General Education Board of New York upon the recommendation of Dr. Frank P. Bachman and the Loans and Scholarships Committee of George Peabody College made it possible financially for the writer to make the present study. The writer is indebted to all of those who had any part in granting the fellowship. She wishes here to express her thanks to Dr. George Harris Webber and to President Beeson of the Georgia State College for Women, for coöperation in the completion of the final study; and also, to her students who have assisted in the clerical work. But she is most deeply indebted to Dr. Joseph Peterson for encouragement, for helpful suggestions, and for understanding criticisms given throughout the progress of the investigation, and especially for those in regard to the application of the tetrad differences criterion.

differences in the intelligence of school children (Peterson, 1925, pp. 111–112). This conclusion suggests that there is little relation between memory and intelligence. In an earlier experimental study of memory (Ebbinghaus, 1885), he limited the problem to an investigation of the conditions which affect memory in any individual. In this study his selection of simple homogeneous material and the elimination of all factors which would affect learning except repetition were apparently based upon the conception that memory is a particular unitary function which could be explained by the then accepted laws of association of learning. The conditions of the experiment limited learning to a stereotyped, rote process. He did show, however, that the measurement of the amount of a given content retained under controlled conditions of learning is a valuable method for testing retention for certain purposes.

Some of the earlier investigations of intelligence included tests of memory, and the general results from these studies indicated some positive relationship between memory and intelligence. Jacobs and Galton in 1887, Bolton in 1891–92 and Kirkpatrick and Bourdon in 1894, through the use of memory tests and other tests of intelligence, found evidence of some relation between memory and intelligence functions (Peterson, 1925, pp. 118–122).

Miss Sharp (1899) published in America the results of an investigation conducted for the purpose of evaluating some of the tests proposed by Binet and Henri as tests of value for 'individual psychology' and for evaluating the theory proposed by them that tests of the higher and more complex thought processes are of more significance for determining individual differences in mental abilities and characteristics than are tests of simpler processes. She used four memory tests, and found that there was practically no tendency for the subjects to maintain consistently their rank on the memory tests when compared with each other nor when compared with their ranks on the other tests.

Winch (1901–02) conducted experiments to investigate the relation between immediate memory and mental ability. He concluded from his results that general ability is accompanied by 'good' memory and that though 'good' memory is usually accompanied by general proficiency, it is not invariably so. He found, also, that memory improves with age, . . . "but principally in so far as increased age itself implies increase of general proficiency" (Winch, 1901–02, p. 134).

Burt (1909) used memory tests in a study of general intelligence. He obtained the following correlations between memory and estimated intelligence; concrete memory and intelligence, .58; abstract memory and intelligence, .48 (Burt, 1909, pp. 176–177).

Simpson (1912) made a study of certain mental abilities to determine what relation they have to each other and to general intelligence. He included in the study three memory tests and a learning test of hieroglyphic forms in pairs which should be grouped with the memory tests. There were twelve other tests of perception, of association, of selective thinking, of sense discrimination and of motor control. The reliability coefficients of the tests are high for all of the tests except the learning of pairs test, but these coefficients are affected by the very wide variability of intelligence within the two groups studied (Kelley, 1916, pp. 223–228). The average correlation of all of the memory tests with all of the other tests was .54; between the memory of passages and all other test results, it was .44. McCall (1916) has shown that the Ebbinghaus completion test used in the study is a good test of intelligence. The high correlation obtained by Simpson between memory and intelligence as measured by this test indicate a decided positive relation between memory and intelligence.

Wyatt (1913–14) made a study of the higher mental processes. He included three memory tests in a group of 15 other tests. The reliability coefficient of the syllables test was .76 and of memory of letter squares was .75. The correlation coefficients of these tests with estimated intelligence were .59 and .18, respectively; that of delayed recall for syllables and intelligence was .74. These correlations are not significant because of the inadequate measure of intelligence, but he obtained a correlation of .65 ± .07 between this test and the completion test. He concluded that his correlations supported Spearman's theory of hierarchical abilities (pp. 109–133).

Kitson (1917) obtained a correlation coefficient of .44 ± .09 between the results of 15 tests of mental ability and school grades for the year 1913–14 and .20 ± .11, for the year 1914–15. These coefficients are low and the latter is unreliable. The only significant correlations obtained were those between memory for logical material seen and heard. In the same year Rosenow (1917) used partial and multiple correlations in the analysis of Kitson's data. He concluded that the logical memory test for material heard was the most significant test in his battery of fifteen tests of mental ability.

Gates (1918) obtained a mean correlation of .82 ± .04 between immediate and delayed recall, and he found that "All tests of immediate and delayed recall show a correlation with teachers' estimate of general intelligence with a central tendency of about .50" (p. 496).

Achilles (1920) studied recognition and recall. She did not study definitely the problem of the relation of memory to intelligence, but she found that the scores made on the tests increased gradually with the age of the children, that there was a tendency for the scores to increase gradually from 4A through 8B and that the younger children in each grade made higher scores. Although no intelligence tests were used, these results indicate in a general way the possibility of a positive relation between memory and intelligence, but they do not prove such a relation.

Miss Lee (1925) studied the problem of the relation of retention to intelligence. She gave group tests of recognition and of recall and forms A and B of the National Intelligence Test to 310 public school children in New York City. The children ranged in age from 7 years and 11 months to 16 years and 3 months and in school achievement from the third through the eighth grade. The following correlations between the recall tests and intelligence were obtained:

> "The average of the correlations between general intelligence and the recall tests of pictures is .52; between general intelligence and the three recall tests of words is .49; between general intelligence and the three recall tests of forms is .30; and between general intelligence and the three recall tests of syllables is .25" (Lee, 1925, p. 13).

These correlations, for such a wide range of abilities represented by the subjects studied, are low but indicate a positive relation between retention and intelligence. The correlations between intelligence and recognition as measured by the different materials and methods are very much lower than those between recall and intelligence. The following statement gives a summary of the means of the coefficients obtained:

> "The average of the three coefficients of correlation between intelligence and (1) recognition of pictures is .34, (2) recognition of words .27, (3) recognition of forms .19, and (4) recognition of syllables .26" (Ibid., 1925, p. 15).

The coefficients are undoubtedly influenced positively by the identical recall elements in the memory tests and the intelligence tests used. The National Intelligence Test includes a code learning test and measures certain memory factors. No reliability measures for the memory tests are given.

Carey (1915) gave six groups of tests to five classes in the senior department of a London County Council elementary school and included ten memory tests in his series of tests. He also included two tests which were calculated to measure the general intelligence factor as conceived by Spearman. By the use of Yule's partial correlation method and a special formula devised by Spearman, he eliminated the influence of *g* ('general intelligence') so that he might determine the relation of the special abilities measured to intelligence. He concluded that there is a small general memory factor apart from intelligence.

Garrett (1928) has recently applied Spearman's tetrad difference formula to the analysis of the intercorrelations between eight memory-learning tests and the Thorndike Intelligence Examination for High School Graduates. He gave his tests to 158 men of sophomore rank or higher who were students at Columbia University in the academic year 1925–26. After partialling out *g*, he found an average correlation of .11, but his probable errors are all around .04 or .05 and he concluded tentatively that there is only a small general memory factor present in the tests. The distribution of tetrad differences from his intercorrelations showed an evidence of overlapping group factors present within the tests, but he accounts for the overlappings as due to similarity of content within the tests and concluded that memory is a general factor.

A critical evaluation of the studies reviewed shows that: (1) No attempt has been made, except in the studies of Simpson, Wyatt, Carey and Garrett to determine the reliability of the memory tests used, and there has been no attempt to ascertain the validity of the memory tests. (2) The measures of intelligence in all of the studies, except that by Garrett, have been inadequate. (3) Since the Binet test, which is considered one of the best instruments for the measurement of intelligence, includes many memory tests, the relations of memory to intelligence should be more definitely investigated.

The present problem is limited to an investigation of the relation of memory to intelligence.

III. *The reliability and the validity of the memory tests.* Before investigating the problem of the relation of memory to intelligence it is necessary to determine the reliability and validity of the memory tests used.

*A. The preliminary study.*—A preliminary study of the reliability of eight widely used memory tests was made in the school year, 1924–25. The tests were given to fifty undergraduate students who were taking the introductory course in General Psychology at the George Peabody College for Teachers.[2] The records of some of the subjects were incom-

[2] The writer was then a student at George Peabody College for Teachers. The work was interrupted by an illness which necessitated an unexpected withdrawal

plete and the reliability coefficients for the tests are based on 44 cases and for some of the tests, on even fewer than this number.

The group tests were given in two periods, each one hour in length, and three weeks later comparable forms of the same tests were given. Copies of the tests will be furnished to anyone who writes the author for them. The first group of the first series of tests was given on April 14, 1925, and the second group of this series, on April 28; the first group of the second series of tests was administered on May 5, and the second group of this series, on May 19. The two forms of the individual tests were given on the intervening days in an order dependent upon the availability of the subjects. The following tests were given according to suggestions adapted from Whipple (1915, pp. 150–203).

1. *The Memory Span for Digits Test.*—The Memory Span for Digits Test consisted of series of digits varying in number from four to twelve. Each series was dictated once and the subjects were asked to write the digits in the order given. The score was the greatest number of digits reproduced in any given series without error.

2. *The Recognition Tests.*—There was a recognition of Words Test and a Recognition of Syllables Test each consisting of 24 items. The items were arranged in serial order on a study sheet which was attached to a test sheet on which the original 24 items were arranged in chance order with 72 items which had not been seen before, making a total of 96 items in the test. The subjects were asked to study the original lists for two minutes and then to turn the sheets and check in the test list those items which were in the original list. The score for the test was one point credit for each item correctly checked, regardless of errors made. The subjects knew nothing of the method of scoring and no statement was made about the number of items either in the study sheet or in the test sheet. The greater number of items in the test list reduced the chances of a pure guess being correct to one in four and made it possible to give credit for each item correctly checked.

3. *The Paired Associates Tests.*—The Recall of Syllables and the Recall of Dates tests each consisted of 16 pairs of items arranged in serial order. The nonsense syllables were arranged in pairs and the subjects were asked to learn the syllables so that when given the 16 syllables in the first column arranged in chance order, they could recall the syllable which had been paired with it in the original list. The Recall of Dates Test consisted of the recall of the several dates of birth of famous persons. The subjects were given a list of sixteen names with the dates of birth. They were asked to learn the dates so that when given the names arranged in chance

---

from school. Because of a very heavy teaching schedule at the Georgia State College for Women, it was impossible to begin the study again until the spring of 1928. At that time she was planning a psychological laboratory and organizing laboratory courses to be introduced into the College curriculum for the first time. This work made the progress of the investigation slower than it would have been otherwise.

order they could give the dates of birth. The test sheet was attached to the study sheet and as soon as time was called the sheets were turned and the dates given. The subjects were asked to draw a circle around any date that was known before the test was given. The study period for each of the tests was six minutes.

4. *The Otis Memory Test.*—Form A of the Otis Memory Test was given with the first series of tests and Form B, with the second series. The instructions given in the Manual for the Otis Group Intelligence Test were used. The questions for each test were printed on mimeographed sheets and included with the other tests given at the time, but there was no possibility of the questions being seen before or while the stories were being read.

5. *The Logical Memory Test.*—A prose selection of comparatively easy subject matter chosen from a book on writing the short story which is not generally known, was used for this test. The selection for each form of the test was cut so that each test consisted of exactly one hundred ideas or phrases. The test was also scored for thought. The score for thought was determined arbitrarily in absolute points. The complete thoughts were assigned a value of ten or fifteen points according to difficulty. Then these complete thoughts were divided into partial thought-units and assigned a value of five or ten or seven or eight points. A score sheet was made and since all of these tests were scored by the writer, the method was uniform and objective even though the weighting was somewhat arbitrary. The test was scored for the number of ideas reproduced in the exact words of the text, one point being given for each idea reproduced. The two sets of scores are treated as two separate tests and are hereafter referred to as the Logical Ideas Test and the Logical Thought Test. The highest number of points which could be made on each of the Logical Ideas Test is one hundred and the highest number of points which could be made on Form I of the Logical Thought Test is ninety. The highest number of points which could be made on Form II of the Logical Thought Test is seventy-five. The time limit for this test was six minutes.

6. *The Recall of Words Test.*—Peterson's method of giving the Recall of Words Test with determination to remember was followed (Peterson, 1916). The subjects were asked to write a list of 20 words as the experimenter wrote them on the board. The words were then pronounced and spelled once by the experimenter and each subject was asked to check her list to see that it was correct. Then the words were erased from the board and the subjects were asked to fold their papers so that the words could not be seen and to write as many of them as they could recall. A similar list of words given in the same manner was used for the second form of the test. One point credit was given for each word correctly recalled.

7. *The Photograph Test.*—The Photograph Test was given as an individual test. There were sixteen photographs in the series, each with a fictitious name. The first name and the family name were given for the women and two initials and the family name were given for the men. The photographs were exposed in a constant series consecutively for six times; then they were exposed in chance order with the name hidden and the subject was asked to write the name of each photograph as it was shown. The rate of exposure was two seconds for each photograph and all the time necessary was given for writing the names. A screen was used to conceal the names and the time was regulated by a muffled metronome. A score of one point was given for each first name, initial or family name correctly recalled. Twenty-four hours later the photographs were exposed in a second random order with the names concealed and the subjects were asked to write those names they could remember.

8. *The Nonsense Syllables Learning Test.*—This test consisted of sixteen nonsense syllables exposed on the Carr Memory Apparatus. The method of giving the test can be understood best from the instructions. They were as follows: "You are to be shown a series of sixteen nonsense syllables. When each syllable is exposed you will spell it aloud. After the series has been exposed one time, you will try to anticipate the syllables and spell them in advance of the exposure. The purpose of the test is to see how quickly you can learn to anticipate the syllables before they are exposed.

"Begin with the first syllable, then learn the second one, then the third one, and so on, adding as many syllables as you can each time until you can anticipate the whole series in advance of the exposure. If you make a mistake or fail to give the syllable on time, the experimenter will spell it for you. Then you are to spell it, and to continue spelling the remainder of the series as you did the first time.

"The exposure will be at the rate of two seconds for each syllable. Try to adapt yourself to the time and give only one syllable in advance of each exposure. Wait for the 'Ready' signal at the beginning of each series."

The score for this test was the number of repetitions necessary for complete learning of the series.

9. *The Army Alpha Intelligence Test.*—The Army Alpha Intelligence Test was given to the subjects by Dr. Peterson. All of the memory tests used in the study were given and scored by the writer, but Dr. Peterson was present when some of the memory tests were given and assisted by timing some of the tests.

Hollingworth (1913), Peterson (1916) and Achilles (1920) have shown that an attitude of 'determination' to learn materially affects the results in learning. In order to keep the influence of attitude constant, an effort was made throughout the instructions to have the subjects take the greatest possible interest in making a high score.

The reliabilities of all of the tests, except that of the Recall of Dates and the Logical Thought tests, were too low even for group measurement of memory and a study of the general relationship between memory and intelligence. Kelley has suggested that a reliability coefficient based upon a single grade range of ability must be .50 or higher for the purpose of group analysis or comparisons (1927, pp. 210 ff.). The 44 subjects who took the tests were all undergraduate college students representing a fairly homogeneous group and Kelley's standard of reliability would be applicable. The reliability coefficient of the Recall of Dates Test is .638 ± .06 and that of the Logical Thought Test is .590 ± .067, but the coefficients of all the other tests are lower than .450. The lowest reliability coefficient, that of the Recall of Words Test, is not four times its probable error. (See Table I.)

The mean of the coefficients for the nine group tests is only .395.

The reliability of the Photograph Test is high in comparison with that of the group tests; the coefficient of the immediate recall test is .552 ± .078 and that of the delayed recall test is .730 ± .051. But the reliability of the Nonsense

TABLE I

RELIABILITY COEFFICIENTS OF THE MEMORY TESTS USED IN THE PRELIMINARY STUDY

| | No. of subjects | $r$ | P.E. |
|---|---|---|---|
| 1. Memory Span for Digits | 45 | .277 | .092 |
| 2. Recognition of Words | 44 | .425 | .083 |
| 3. Recognition of Syllables | 44 | .348 | .089 |
| 4. Recall of Dates | 44 | .638 | .060 |
| 5. Paired Associates Syllables | 44 | .384 | .087 |
| 6. Otis Memory A and B | 44 | .357 | .090 |
| 7. Logical Ideas | 44 | .364 | .088 |
| 8. Logical Thought | 44 | .590 | .067 |
| 9. Recall of Words | 44 | .174 | .097 |
| 10. Photograph Test—Immediate Recall | 36 | .552 | .078 |
| 11. Photograph Test—Delayed Recall | 36 | .739 | .051 |
| 12. Nonsense Syllables Learning Test | 36 | .390 | .099 |

Note: *All reliability coefficients of the above tests are based on repetitions of a second form of a comparable test.*

Syllables Learning Test is very low, being only .390 ± .099. When an analysis of the results of the preliminary study was made in 1924–25 and the tests were found to be too unreliable for an accurate measure of memory it was thought that the tests would be increased in length on the basis of the Spearman-Brown prophecy formula. But Holzinger (1923) showed that when the formula was applied to the Terman group test of intelligence, the predictions were lower except for the first four or five additional tests than the formula would indicate. Holzinger and Clayton (1925) showed that except for the 'best fit' values of $r_{xx}$ and for simple homogeneous material the formula cannot be applied. Lanier (1927) showed that when the formula was applied to the Otis intelligence test and to his Tapping Test II that it did check up rather closely, but that when it was applied to 10 other tests of musical and mechanical abilities there was

great variation between the actual and predicted reliabilities. From these studies it is evident that there are other factors besides the length of a test which affect its reliability. Because of this fact the tests were modified empirically from an examination of the mean scores and the distributions of scores.

All of the memory tests used in the present study, except the Nonsense Syllables Learning Test are tests of retained members. The subject matter of a memory test of this type is studied for a given length of time or for a given number of repetitions and the amount of content reproduced in test score units is the measure of memory ability (Whipple, 1915, pp. 151). An analysis of the mean scores made on the tests showed that there are two problems involved in determining the length of a memory test. The first is to determine the amount of content which should be included in the test; the second, is to determine the length of study period which will give the best distribution of ability to learn the given content. A memory test of relatively easy subject matter which measures ability to learn a simple pattern series should be lengthened by increasing the number of items to be studied in a given length of time; a memory test of difficult material should be lengthened by increasing the time limit for study.

The scores for all of the tests were higher on Form II of the test than on Form I. An examination of individual scores made on the two series of tests revealed discrepancies in the scores that were evidently due to a lack of understanding of the first tests and to practice effects. In order to make the conditions of learnings for the two series of tests more comparable a fore-exercise was added to each test used in the final study except the Recall of Words Test. Kelley has suggested that to give a fore-exercise is one of the three conditions which make for the reliability of a test (1923, pp. 203). The instructions for each test given to Georgia students were made more definite.

The reliability coefficient for the Memory Span of Digits Test was only $.277 \pm .092$, and since it was not possible to take time to give a sufficient number of series of digits to reduce the accidental factors affecting the scores, this test

was omitted from the final study. The Paired Associates Syllables Test had a very low mean score and the test was difficult and so disliked by the students that it was omitted from the final study. It is unfortunate that these omissions were necessary, but the tests will be studied further in future investigations.

The means of both forms of the Recognition of Syllables Test were slightly higher than fifty percent of the total number of items in the tests, but there were so few high scores made that an extension of the time limit for study seemed necessary (see Table II) and the time was extended from

TABLE II

MEAN SCORES OF THE MEMORY TESTS USED IN THE PRELIMINARY STUDY

| Test | No. items in test | Time limit | Means | Sigmas |
|---|---|---|---|---|
| 1. Memory Span for Digits | I 4–12 | — | 8.17 | 1.38 |
| 2. Memory Span for Digits | II 4–12 | — | 9.35 | 1.48 |
| 3. Recognition of Words | I 24 | 2 min. | 16.32 | 2.95 |
| 4. Recognition of Words | II 24 | 2 " | 18.98 | 3.84 |
| 5. Recognition of Syllables | I 24 | 2 " | 15.14 | 3.55 |
| 6. Recognition of Syllables | II 24 | 2 " | 16.52 | 3.43 |
| 7. Recall of Dates | I 16 | 6 " | 8.23 | 3.13 |
| 8. Recall of Dates | II 16 | 6 " | 10.5 | 3.16 |
| 9. Paired Associates Syllables | I 16 | 6 " | 4.52 | 3.13 |
| 10. Paired Associates Syllables | II 16 | 6 " | 5.96 | 2.82 |
| 11. Otis Memory Test | A 30 | — | 21.89 | 2.81 |
| 12. Otis Memory Test | B 30 | — | 23.80 | 2.43 |
| 13. Logical Ideas | I 100 | 3 min. | 28.32 | 14.02 |
| 14. Logical Ideas | II 100 | 3 " | 23.96 | 7.96 |
| 15. Logical Thought | I 90 | 3 " | 34.09 | 16.94 |
| 16. Logical Thought | II 70 | 3 " | 55.64 | 22.53 |
| 17. Recall of Words | I 20 | — | 13.43 | 2.45 |
| 18. Recall of Words | II 20 | — | 15.75 | 2.61 |
| 19. Photograph Test—Immediate Recall | I 39 | 6 repetitions | 26.17 | 7.99 |
| 20. Photograph Test—Immediate Recall | II 39 | 6 repetitions | 27.39 | 7.19 |
| 21. Photograph Test—Delayed Recall | I 39 | — | 21.83 | 8.52 |
| 22. Photograph Test—Delayed Recall | II 39 | — | 20.72 | 8.85 |

two to three minutes. In order to compare the results of the Recognition of Words Test with those of the Recognition of Syllables Test, the time of the former was extended to three minutes also. The extension of the study period for this test was a mistake, however. The longer time limit made it possible for so many subjects to learn the words completely that the scores became too much bunched in the upper range.

The average of the mean scores on each of the two forms of the Logical Ideas Test was only 26.15 with an average sigma deviation of 10.99. The highest possible score which could have been made on this test was 100. The average of the mean scores made on the two forms of the Logical Thought was only 45.87, with a mean sigma of 11.73. The highest possible score which could have been made on Form I of this test was 90, and on Form II it was 70. These low scores indicated the necessity of increasing the study period in this test. The time limit was extended from three minutes to six. The results from this increase in time for study were more significant than those from any other change made in the methods of giving the tests, as will be shown in the analysis of the reliability of the final tests. No change in the time limit for study was made in the Recall of Dates Test. Although the reliability of the Otis Memory Test was low, no change was made in the method of giving this test. The average of the means of the scores made on the two forms of the Recall of Words Test was 14.59 and there were only twenty items in the original test. The number of items in this test was increased from twenty to twenty-four.

*B. The reliability of the final tests.*—The tests included in the final study are: Recognition of Words, Recognition of Syllables, Recall of Dates, Otis Memory, Logical Ideas, Logical Thought, and Recall of Words. The Otis Self-Administering Higher Examination and the Army Alpha intelligence tests, forms A and 9 respectively, were given as measures of intelligence.

These tests were given to 225 (white) students in the freshman class at the Georgia State College for Women in the spring semester of 1928. The students were taking History of Education under the writer's instructions and were grouped for the work in four class-sections. Only 200 records were complete and all correlations are based on this number. The chronological ages of the subjects ranged from 15 years and nine months to 25 years and 3 months. The mean chronological age was 17 years and 11 months with a sigma deviation of 1 year and 2 months.

The Army Alpha intelligence test was given to all of the groups during the month of March before the testing program was actually begun. The other tests were given in three series on three succeeding days of the same week. The first series of tests was given to the first group of students on May 9, and the last series was given to the last group on May 25. The Recognition of Words, the Recognition of Syllables, and the Recall of Dates and the Logical Memory tests were given in the one-hour class period of the first day the testing program was begun with any section; the Otis Self-Administering intelligence test was given on the second day; the Recall of Words and the Otis Memory tests and a Memory-of-Sentences test were given in the last series on the third day. One week later comparable forms of the tests were given in the same serial order. The Memory-of-Sentences Test is not included in the study because a satisfactory method of scoring has not yet been worked out.

The total correlations were computed by the Pearson product moment formula, the partial and multiple correlations, by Yule's formulæ for partial and multiple correlations (Yule, 1912) and the probable errors for the total, the partial and the multiple correlations, by Kelley's formulæ (Kelley, 1923, pp. 200–310). The correlations were obtained from the average scores made on the two forms of the memory tests in all cases except that of the Memory of Dates. The sums of the scores made on the two forms were used for this test.

The total correlations from the raw scores which gave any indication of being non-linear were tested for linearity by Blakeman's short test for linearity (Holzinger, 1928, pp. 183–184). All of the correlations involving the tests of Recognition of Words and of Recall of Words were thus tested. The relation between the Army Alpha and the Recall of Words and that between the Recognition of Syllables and the Recall of Words are slightly non-linear, but those between all the other tests are linear. The relation between the Recall of Words and the Otis intelligence tests is linear and therefore the correlations involving the Army Alpha test are eliminated from some of the more important analyses.

The reliability of the Otis Self-Administering and the Army Alpha intelligence tests was determined by correlating the odds with the evens. The reliability of the Otis Self-Administering is .830 ± .015. This coefficient is decidedly lower than .921 which is given by Otis for a group of 128 subjects ranging from the seventh through the twelfth grades (1922, pp. 12). The reliability obtained by correlating parts of the same test by the odds with the evens method is too high because subjects are more likely to have the same attitude toward a test given at one sitting than if comparable forms of the same test are given at different sittings. The reliability of the Army Alpha for the group of subjects used in the present study is .849 ± .013.

There were two contributing factors operating to cause differences in the reliability of the memory tests used in the preliminary and those used in the final study. The fore-exercises were added to the second tests and the time for study or the number of items in a number of tests was increased. There was only a slight increase in the reliability of the Recognition of Words Test, but an increase of .211 points in the reliability of the Recognition of Syllables Test. (See Table III.) If the Spearman-Brown prediction formula

TABLE III

THE INCREASE IN RELIABILITY RESULTING FROM CHANGES MADE IN GIVING THE MEMORY TESTS

| | Peabody group | | G.S.C.W. group | | Expected r's | Increase in r's |
|---|---|---|---|---|---|---|
| | r | P.E. | r | P.E. | r | P.E. |
| 1. Recognition of Words | .425 | .092 | .478 | .037 | .526 | .053 |
| 2. Recognition of Syllables | .348 | .089 | .559 | .032 | .445 | .211 |
| 3. Recall of Dates | .638 | .060 | .615 | .030 | — | .023 |
| 4. Otis Memory | .357 | .090 | .343 | .042 | — | .014 |
| 5. Logical Ideas | .364 | .088 | .583 | .032 | .534 | .219 |
| 6. Logical Thought | .590 | .067 | .686 | .096 | .742 | .096 |
| 7. Recall of Words | .174 | .097 | .472 | .037 | — | .298 |

is applied to the initial coefficient of the Recognition of Words Test, the expected reliability coefficient resulting from

increasing the time would be .526; the actual coefficient obtained is .478 ± .037. The expected coefficient of the Recognition of Syllables Test is .445; the obtained coefficient is .559 ± .032. It is not possible to determine objectively the influence of the fore-exercise upon the reliability of the tests. But if the conditions of giving the tests had been the same the increase in the time limit for the Recognition of Words Test would not have given the expected increase in reliability. The increase in the study period of this test enabled many subjects to make a perfect score on the test and although their scores are practically the same for both forms, as was stated above, they are much too bunched in the upper limit to differentiate well among the several subjects. The nonsense syllables had never been seen by the subjects before; past memory experiences could have little influence in facilitating the learning. Therefore, the test was more difficult and the longer study period gave a more accurate measure of ability to learn this type of material.

The Recall of Dates Test and the Otis Memory Test were given in the final study just as in the preliminary study. The reliability coefficients of both of these tests were slightly lower for the two hundred freshmen subjects than for the forty-four Peabody students. These results agree with Kelley's statement that the comparison of the reliability coefficients from different groups is of little value unless the homogeneity of the groups is known (1923, pp. 223).

The increase in the study period of the logical memory test resulted in a very significant increase in its reliability. It was pointed out above that the test was scored both for ideas and for thought reproduced. The two factors which would cause an increase in the reliability of the tests are the addition of the fore-exercise and the increase in the time limit for study. On the basis of the Spearman-Brown prediction formula the expected reliability coefficient of the Logical Ideas Test would be .534; the actual coefficient obtained is .583 ± .032; the expected coefficient of the Logical Thought Test would be .742, but the obtained coefficient is only .686 ± .096. Because the other factors

influencing the reliability of the tests cannot be isolated objectively, the application of the prediction formula to these results can be only suggestive. But they seem to indicate that the formula applies rather well to the logical memory tests and not very well to the simple recognition memory tests. The reliability coefficient of the Recall of Words Test was lower than that of any other test used in the final study except that of the Otis Memory Test. The increase in the number of items used in the test resulted, however, in an increase in reliability of .30 points over the first coefficient.

The reliability coefficient for a composite of all of the memory tests obtained from the mean centile scores of all the tests is $.739 \pm .021$. This reliability coefficient was obtained by correlating the mean of the centile scores made by the several subjects on the first form of all the memory tests with the mean of the centile scores made on the second form of the tests. (For an explanation of the method of computing the centile scores, see the discussion of the validity of the tests.)

*C. The validity of the memory tests used in the final study.* —There is no reliable criterion available for determining the validity of memory tests. If a reliable and valid test of memory were available that would be the best criterion. Since a reliable criterion was not available, it was necessary to use the indirect method of correlating each memory test with a composite of all the other tests. The composite scores from which the validity coefficient were compiled were the means of the individual centile scores made on Forms I and II of the memory tests.

The raw scores of all of the tests were converted into centile scores in terms of sigma units on the basis of Table VI in Rugg's *Statistical Methods Applied to Education.* This table is based upon the area of the probability curve which assumes the base line to be broken off at plus or minus 3 sigmas. The centile scores are stated in sigma units on a scale from 0 to 100 in which the zero point is set at minus 3 sigmas, the 50 centile is set at the mean and the 100 point

at plus 3 sigmas. Rugg uses the term 'percentile score' for this sigma unit of measure, but the term 'centile score' is used here to distinguish it from the percentile score which generally means . . . "a value of the variable below which

TABLE IV

CORRELATION COEFFICIENTS SHOWING THE VALIDITY OF THE TESTS AND THE MEAN OF THE TOTAL COEFFICIENTS OF EACH TEST CORRELATED WITH EACH OF THE OTHER TESTS

| Tests | Validity coefficients | P.E. | Mean of total correlations of each test with all other tests |
|---|---|---|---|
| 1. Recognition of Words | .410 | .039 | .279 |
| 2. Recognition of Syllables | .403 | .040 | .282 |
| 3. Recall of Dates | .446 | .038 | .300 |
| 4. Otis Memory Test | .374 | .041 | .257 |
| 5. Logical Ideas | .658 | .027 | .427 |
| 6. Logical Thought | .659 | .027 | .411 |
| 7. Recall of Words | .369 | .041 | .243 |

a given percent of the frequencies lie" (Holzinger, 1928, pp. 127), or the percentage of subjects making scores below a given score. To facilitate the computation a table for converting the raw scores into centile scores was constructed for each test. The intervals of the table were in .06 standard deviation units above and below the mean. The centile scores were read from the table by comparing the raw scores with the corresponding points on the scale.

After the raw scores had been reduced to centile scores seven sets of composite scores were obtained. The validity coefficients were obtained by correlating the means of the individual centile scores made on Forms I and II of each test with the composite scores which did not include the test in question. For example, the validity coefficient for the Recognition of Words Test was obtained by correlating the means of the centile scores for Recognition of Words Forms I and II with the means of the average centile scores made on Forms I and II of all the other memory tests. The correlation coefficients showing the validity of the tests as measured by this indirect method are given in Table IV.

The Otis Memory and the Recall of Words tests are of doubtful validity even for group comparisons; and the

validity of the Recognition of Words, of the Recognition of Syllables, and of the Recall of Dates tests is low when evaluated by this criterion. The total correlation coefficients between the two recognition tests is higher than that between either and the composite of all the other memory tests. Achilles (1920) and Lee (1925) have shown that the correlation between scores made on recognition and on recall tests of memory are low. The total intercorrelations between the recognition and the recall tests used in the present study substantiate these results. Therefore recognition tests would not be expected to correlate highly with a composite in which most of the tests are recall tests.

The validity coefficient of the Otis Memory Test is very low. The partial correlations show that when intelligence as measured by the Otis Self-Administering and the Army Alpha tests is held constant, the mean of the coefficients between the Otis Memory and the other tests is only .065. Since the reliability of the test is low, any conclusions in regard to it must be tentative, but it seems to measure the same type of ability as that measured by the intelligence tests. The mean of the coefficients between this test and the two intelligence tests is .539. The Army Alpha and the Otis Intelligence tests both measure certain memory functions and it is possible that the relation between the Otis Memory Test and these tests is due to the common memory factors in the tests, but it is also possible that the Otis Memory Test measures general intelligence primarily. The responses made to this test do not involve the visual and motor cues which were utilized in reading and studying the other tests. The necessity of answering questions does not make possible simple serial or consecutive reinstatement of the previous act with each preceding partial response as a stimulus facilitating the following partial response. The subject does not see the questions until after he has heard the story and there is little opportunity to organize and group material for specific reproduction through the utilization of specific cues. The test may measure a high type of ability to give a delayed multiple-choice language response. It is possible that the

test measures an ability to give a delayed reaction which is expressed only in language and thinking and that this ability is characteristic of intelligence. This interpretation is merely suggestive, however, because the low intercorrelations may be due entirely to the unreliability of the test.

The validity coefficients of the Logical Ideas and the Logical Thought tests are practically the same, but the coefficient of alienation between the two tests is .50. The tests measure much the same, though probably not identical phases of memory. The logical memory tests have higher validity coefficients than any of the other memory tests. The intercorrelations between these tests and the intelligence tests are higher than those of any of the other memory tests used, but the intercorrelations between these tests and the

TABLE V

THE INTERCORRELATIONS BETWEEN SEVEN MEMORY TESTS WITH TWO TESTS OF INTELLIGENCE AND WITH EACH OTHER

| Tests | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Otis Self-Administering Intelligence | *.830*<br>*.015* | .831<br>.015 | .216<br>.046 | .211<br>.046 | .381<br>.041 | .529<br>.034 | .561<br>.034 | .639<br>.028 | .179<br>.046 |
| 2. Army Alpha | | *.849*<br>*.013* | .149<br>.037 | .204<br>.046 | .377<br>.041 | .548<br>.033 | .542<br>.034 | .633<br>.025 | .164<br>.046 |
| 3. Recognition of Words | | | *.478*<br>*.037* | .457<br>.038 | .182<br>.046 | .206<br>.046 | .301<br>.043 | .303<br>.043 | .227<br>.045 |
| 4. Recognition of Syllables | | | | *.559*<br>*.032* | .208<br>.046 | .211<br>.026 | .362<br>.042 | .344<br>.042 | .110<br>.047 |
| 5. Recall of Dates | | | | | *.615*<br>*.030* | .198<br>.046 | .357<br>.042 | .365<br>.041 | .489<br>.036 |
| 6. Otis Memory | | | | | | *.343*<br>*.042* | .379<br>.041 | .409<br>.040 | .136<br>.047 |
| 7. Logical Ideas | | | | | | | *.583*<br>*.032* | .856<br>.013 | .308<br>.043 |
| 8. Logical Thought | | | | | | | | *.686*<br>*.025* | .189<br>.046 |
| 9. Recall of Words | | | | | | | | | *.472*<br>*.037* |
| Mean of each memory test with all other memory tests | | | .279 | .282 | .300 | .257 | .427 | .419 | .243 |

other memory tests are least affected when intelligence is held constant.

The Recall of Words Test is the least valid of all the tests. This may be due to the fact that it gives so little opportunity for learning. The test must measure some distinct memory function however, because the coefficients of relationship between this test and the Recall of Dates and the Logical Ideas tests are significantly positive.

IV. *The relations of memory functions to intelligence.*—The correlation between the Recognition of Words and the

TABLE VI

PARTIAL CORRELATIONS BETWEEN THE MEMORY TESTS WITH THE INTELLIGENCE TESTS HELD CONSTANT

Index to variables:

1—Otis Self-Administering
2—Army Alpha
3—Recognition of Words
4—Recognition of Syllables
5—Recall of Dates
6—Otis Memory
7—Logical Ideas
8—Logical Thought
9—Recall of Words

| Total coefficients | | | Tests and those held constant | | | Tests and those held constant | | | Tests and those held constant | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Tests | *r* | P.E. | Tests | *r* | P.E. | Tests | *r* | P.E. | Tests | *r* | P.E. |
| 34 | .457 | .038 | 34.1 | .431 | .039 | 34.2 | .439 | .038 | 34.12 | .434 | .039 |
| 35 | .182 | .046 | 35.1 | .111 | .047 | 35.2 | .126 | .047 | 35.12 | .118 | .047 |
| 36 | .206 | .046 | 36.1 | .111 | .047 | 36.2 | .149 | .047 | 36.12 | .128 | .047 |
| 37 | .301 | .043 | 37.1 | .225 | .045 | 37.2 | .265 | .044 | 37.12 | .237 | .045 |
| 38 | .303 | .043 | 38.1 | .219 | .045 | 38.2 | .279 | .044 | 38.12 | .239 | .045 |
| 39 | .227 | .045 | 39.1 | .196 | .046 | 39.2 | .208 | .046 | 39.12 | .197 | .046 |
| 45 | .208 | .046 | 45.1 | .142 | .047 | 45.2 | .141 | .047 | 45.12 | .137 | .047 |
| 46 | .211 | .046 | 46.1 | .119 | .047 | 46.2 | .117 | .047 | 46.12 | .110 | .047 |
| 47 | .362 | .042 | 47.1 | .302 | .043 | 47.2 | .300 | .043 | 47.12 | .297 | .045 |
| 48 | .344 | .042 | 48.1 | .277 | .044 | 48.2 | .279 | .044 | 48.12 | .273 | .044 |
| 49 | .110 | .047 | 49.1 | .075 | .047 | 49.2 | .120 | .047 | 49.12 | .074 | .047 |
| 56 | .198 | .046 | 56.1 | −.005 | .048 | 56.2 | −.012 | .048 | 56.12 | −.003 | .048 |
| 57 | .357 | .042 | 57.1 | .196 | .046 | 57.2 | .195 | .046 | 57.12 | .079 | .047 |
| 58 | .365 | .041 | 58.1 | .171 | .046 | 58.2 | .158 | .047 | 58.12 | .148 | .047 |
| 59 | .489 | .036 | 59.1 | .463 | .038 | 59.2 | .467 | .037 | 59.12 | .463 | .037 |
| 67 | .379 | .041 | 67.1 | .047 | .048 | 67.2 | .119 | .047 | 67.12 | .082 | .047 |
| 68 | .409 | .040 | 68.1 | .109 | .047 | 78.2 | .053 | .048 | 68.12 | .056 | .048 |
| 69 | .136 | .047 | 69.1 | .049 | .048 | 69.2 | .056 | .048 | 69.12 | .044 | .048 |
| 78 | .856 | .013 | 78.1 | .779 | .019 | 78.2 | .792 | .018 | 78.12 | .774 | .019 |
| 79 | .308 | .043 | 79.1 | .255 | .045 | 79.2 | .264 | .044 | 79.12 | .255 | .045 |
| 89 | .189 | .046 | 89.1 | .099 | .047 | 89.2 | .108 | .047 | 89.12 | .095 | .047 |

Recognition of Syllables tests is almost twice as great as the mean of the coefficients of relationship between either of the recognition tests with all of the recall tests. (See Tables V and VII.) When intelligence as measured by the Otis and

TABLE VII

THE MEANS OF THE CORRELATIONS BETWEEN EACH TEST WITH THE RECOGNITION TESTS AND WITH THE RECALL TESTS WHEN INTELLIGENCE IS HELD CONSTANT

| Tests | Means of Correlations with Recognition Tests | Means of Correlations with the Recall Tests |
|---|---|---|
| 1. Recognition of Words | .457 | .244 |
| 2. With intelligence constant | .439 | .184 |
| 3. Recognition of Syllables | .457 | .243 |
| 4. With intelligence constant | .439 | .178 |
| 5. Recall of Dates | .195 | .352 |
| 6. With intelligence constant | .128 | .164 |
| 7. Otis Memory | .209 | .281 |
| 8. With intelligence constant | .119 | .037 |
| 9. Logical Ideas | .332 | .475 |
| 10. With intelligence constant | .267 | .298 |
| 11. Logical Thought | .324 | .455 |
| 12. With intelligence constant | .256 | .268 |
| 13. Recall of Words | .169 | .274 |
| 14. With intelligence constant | .136 | .221 |

the Army Alpha tests is held constant, the coefficient of relationship between the two recognition tests is only slightly reduced, but the correlations between the recognition and the recall tests are decidedly decreased when intelligence is controlled. When the influence of intelligence is controlled, however, there is still a mean positive relationship of .184 between Recognition of Words and the recall tests, and, of .178 between the Recognition of Syllables and the recall tests. This mean residual relationship is in each case reliable.

The mean of the coefficients between the Recall of Dates and the two recognition tests is only .195 and this is reduced to .128 when intelligence is controlled; the mean of the coefficients between the Recall of Dates Test and the recall tests is .352, but this mean coefficient is reduced to .164 when the influence of intelligence is held constant. However, the correlation between this and the Recall of Words Test with

the influence of intelligence controlled is .463 ± .037. This high coefficient indicates that these two tests have some factors in common which are not measured by the intelligence tests.

It was pointed out in the analysis of the validity of the Otis Memory test that the mean of the coefficients between this test and all of the other tests when intelligence is held constant is .065. This zero relationship between the Otis Memory Test and the other tests when intelligence is controlled may be due to the unreliability of the test, but it seems to indicate that this test measures the same mental functions as those measured by the intelligence tests.

The residual relationship between the memory functions measured by the logical memory tests is higher when the influence of intelligence is controlled than that of any other memory test. The mean of the coefficients between the Logical Ideas and the two recognition tests is still .267 when intelligence is held constant; and that between Logical Thought and the two recognition tests is still .256. (See Table VII.) The mean of the coefficients between the Logical Ideas Test and the recall tests without the influence of intelligence is .298; and that between the Logical Thought and the recall tests, .268. This is very much greater than the corresponding mean of the coefficients of any other memory test with the influence of intelligence controlled. The high residual relationship between the logical memory tests and the recall tests is partly due, however, to the high intercorrelation between the logical memory tests themselves and it was pointed out above that they measure factors that are to a great extent identical.

The mean of the coefficients between the Recall of Words and the recognition tests is very low, but this relationship is only slightly influenced by the intelligence factors as measured by the two intelligence tests. The mean of the coefficients between this test and the other recall test is .282 and this mean is still .214 with intelligence constant. (See Table VII.) It was previously stated that there is a high positive relationship between the factors measured by the Recall of Dates

Test and those measured by this test. There is also a high correlation between this test and the Logical Ideas Test. When intelligence is held constant there is little change in either of these coefficients.

The mean of the correlations of each memory test with all others, when intelligence as measured by both the Otis and the Army Alpha intelligence tests is controlled, is still .202 with a mean probable error of .044. This reliable mean residual relationship which is independent of the mental functions measured by the intelligence tests indicates the presence in the memory tests of memory factors that are not measured by the intelligence tests. This reliable residual relationship is due, however, to the high intercorrelations between the two recognition tests, between these two tests and the logical memory tests, and between the Recall of Words and the Logical Ideas and the Recall of Dates tests. The recall of meaningful material which involves learning seems to involve mental functions that are very closely related to intelligence.

When the Otis Self-Administering intelligence test as the dependent variable is correlated with two recognition tests as independent variables the multiple correlation coefficient is only .337 ± .042. When the Recall of Words Test, which is the simplest of the recall tests and involves practically no learning, is added as a third independent variable, the coefficient becomes only .364 ± .041. But the multiple correlation between the Otis Self-Administering on the one hand and the Recall of Dates and the Logical Thought tests on the other as dependent variables is .659 ± .027. When the Logical Ideas Test is added as a third independent variable, the coefficient becomes .695 ± .025. When the last three tests are correlated with the Army Alpha as the dependent variable the multiple correlations are practically the same. (See Table VIII.) The multiple correlation of the Recall of Dates and the Logical Thought tests combined with the Otis Self-Administering Test is higher than the coefficient between either of the intelligence tests and the composite of centile scores made on all of the memory tests. The correlation between the composite of the centile

scores made on all of the memory tests and the Otis Self-Administering Test is .581 ± .031 and between the composite of the memory tests and Army Alpha it is .501 ± .036. Garrett has recently (1928, pp. 605) obtained a multiple

TABLE VIII

MULTIPLE CORRELATIONS SHOWING THE RELATION BETWEEN MEMORY AS MEASURED BY THE BEST WEIGHTING OF THE SIMPLE AND OF THE MEANINGFUL MEMORY TESTS AND THE OTIS AND THE ARMY ALPHA INTELLIGENCE TESTS

Index to variables:

| | |
|---|---|
| 1—Otis Self-Administering | 5—Recall of Dates |
| 2—Army Alpha | 7—Logical Ideas |
| 3—Recognition of Words | 8—Logical Thought |
| 4—Recognition of Syllables | 9—Recall of Words |

| Tests | Correlation | P.E. |
|---|---|---|
| 1.34 | .337 | .042 |
| 1.349 | .364 | .041 |
| 1.57 | .611 | .030 |
| 1.58 | .659 | .027 |
| 1.578 | .695 | .025 |
| 2.57 | .577 | .031 |
| 2.58 | .652 | .027 |
| 2.578 | .649 | .028 |

correlation of .53 between the Thorndike Intelligence Test and eight memory-learning tests.

V. *Specific analysis of the relation of memory to intelligence based upon Spearman's criterion of tetrad differences.*—Spearman (1904, 1914, 1927) has developed the criterion of tetrad differences which may be used to determine the presence or absence of a general factor in a group of correlated variables. Since the formula for the probable error of a distribution of tetrad differences has been made available recently, it is possible to apply the tetrad difference method to our results, in the endeavor to ascertain whether memory is only an aspect of general intelligence or whether there may be general memory factors that are independent of intelligence. The reader who is unacquainted with the Spearman technique is referred to the Appendix in the *Abilities of Man.*

The criterion of tetrad differences has been devised by Spearman to prove his theory, that when the members of any pair of mental traits are correlated to any extent with each other, they are to that extent dependent upon a common factor which he calls general intelligence or *g*. Since applying

the tetrad-differences technique to many groups of measurements of mental functions, Spearman has concluded that his original two-factor theory must be modified because of the probable existence of two other general intelligence factors, which he considers aspects of *g*, and special group factors. From a similar analysis of studies of memory he has concluded that memory involving pure retentivity is a complex function of group factors; and that the correlations between pure retentivity and intelligence are low, but the correlations between memory tests which involve an eduction of relationships and intelligence are positive and significant.

A tetrad consists of any four coefficients which form a rectangle in a correlation table. The formula for determining the tetrad difference of a given tetrad is $r_{ap} \cdot r_{bq} - r_{bp} \cdot r_{aq} = 0$ (1927, Appendix, pp. viii). But the formula for the tetrad differences of any four correlations which form a tetrad in a correlation table could be true only for 'true correlations,' and since the correlations between scores made on tests of mental abilities are obtained from mere samples of population the tetrad differences will not actually equal zero; but if the theory is true, they will distribute themselves from zero in chance order (Spearman, 1927, pp. 140–141). In order to apply the formula to test data it is necessary to compare the obtained distribution of tetrad differences with the theoretical distribution expected from chance errors of sampling. A general evaluation may be made by comparing the median of the obtained tetrad differences with the theoretical probable error of these differences computed from the intercorrelations of the variables involved. If the obtained median is less than or equal to the theoretical probable error of the tetrad differences, then the obtained distribution agrees with the theoretical chance distribution and the intercorrelations must be explained on the basis of the common general factor, *g*, and specific factors. If the obtained median is greater than the theoretical probable error, the intercorrelations must be explained as due to the overlapping of group factors, unless the correlations are spurious or are obtained from scores made on tests which measure the same factor.

Kelley (1928, pp. 14) has offered two fundamental criticisms of the tetrad criterion as interpreted by Spearman. The first is that since some of the chance errors in the tetrads may be correlated with each other, it is unreasonable to assume that the tetrad differences will form a normal frequency distribution in accordance with the laws of chance. The second criticism is that all that Spearman's criterion can show is whether or not one factor is or is not sufficient to explain a given situation. By mathematical proof he has extended the conception of the tetrad criterion to show the necessity of considering many possible combinations of general and specific or group factors. Through the application of his own iteration method he has found six general and nine special factors which are specific mental traits.

In the present study there are two pairs of tests whose intercorrelations with each other and with the other tests seem to indicate that they are at least partly duplicate measures of the same mental functions. They are the Otis Self-Administering and the Army Alpha intelligence tests and the Logical Ideas and the Logical Thought memory tests. The Logical Ideas and the Logical Thought tests are different scores for the same test. When they are correlated the coefficient of alienation is .50 and the differences in the total correlations involving the two tests show that they are not exactly identical measures. But because the two pairs of tests mentioned duplicate each other to some extent, the Army Alpha and the Logical Ideas tests are eliminated from the table of intercorrelations from which the tetrad differences were obtained. As stated above, there is one non-linear regression with the Army Alpha test, and the Logical Thought Test is more reliable than the Logical Ideas.

The intercorrelations between the variables from which the tetrad differences are obtained are given in Table V, but as previously stated, the Army Alpha and the Logical Ideas tests are eliminated from this analysis. There are 105 tetrad differences ($3 \times 7^c4$) for the seven variables considered. The obtained median for these tetrad differences without regard to sign is .0395; the mean is .0100; the obtained

standard deviation is .0874 and the probable error is .0590. The theoretical probable error for the expected chance frequency of tetrads is .0238. This probable error was derived by Spearman's formula 15 (1927, Appendix, pp. vii). The formula is

$$\text{P.E.} \frac{1.349}{\text{N}} \sqrt{r^2(1 - r)^2 - s^2}.$$

The $r$ in the formula is the mean of all of the coefficients from which the tetrads are obtained and $s^2$ is the mean squared deviation of all the $r$'s from their mean. This formula gives a probable error that is sometimes too large, but it will give a value that is approximately the same as that obtained by the use of his later formula 16A (1927, Appendix, pp. xi). The probable error computed by formula 16A is .0213. The median of the obtained tetrad differences is .0395 which is much larger than the theoretical probable error. This indicates the presence of overlapping group factors, and not a single factor with specific factors as the cause of the intercorrelations between the variables. The Otis Intelligence Test probably measures not a discrete particular aspect of intelligence, but a number of complex mental functions. The large difference between the obtained median of the tetrads and the theoretical probable error indicates that the intercorrelations between the memory tests and the intelligence test cannot be explained as due to a general intelligence factor and a specific memory factor.

According to the six formulæ given by Kelley (1928, pp. 47), there are twice as many tetrad differences as are calculated and those that are not calculated are the same as those that are, but they have signs that are opposite. The 210 tetrad differences from the seven variables considered are plotted in Fig. 1 in a histogram together with a curve showing the chance distribution expected from the theoretical probable error. A comparison of these frequency distributions shows that there are 46 tetrad differences that are greater than $\pm$ 3 sigmas from the theoretical mean. These high tetrad differences cannot be explained as due to chance errors of sampling. The analysis of partial correlations showed that the Otis

Memory Test probably measures the same mental functions as those measured by the Otis Intelligence Test; the reliability of the test is not satisfactory, and therefore the high tetrad differences caused by this test should not be considered significant. There are, however, 32 remaining tetrad differences beyond ± 3 sigmas which cannot be explained as due to chance errors of sampling. The intercorrelations between

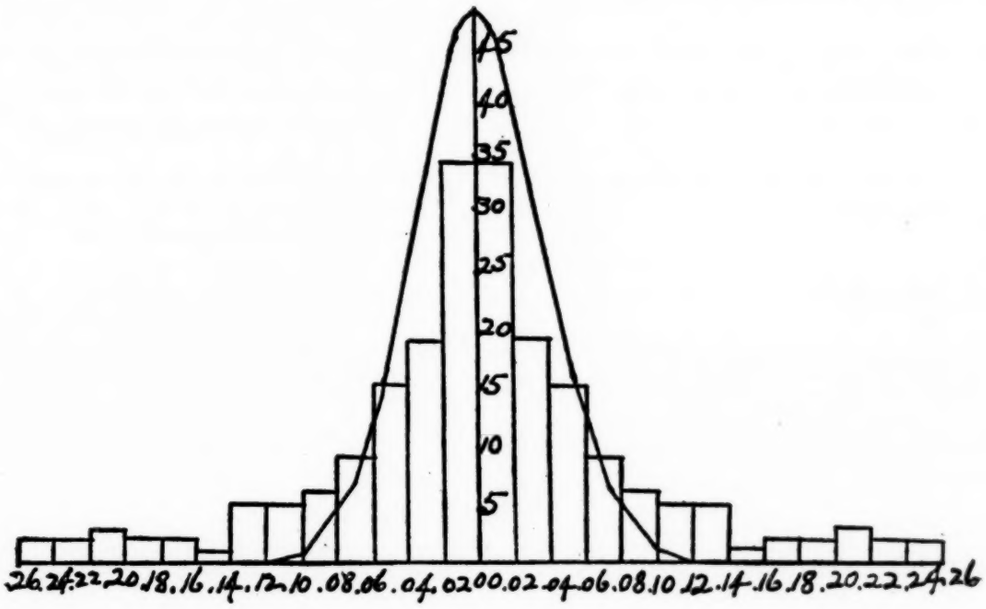



FIG. I. Frequency distribution of 210 tetrad differences compared with the distribution expected from chance errors of sampling.

the variables from which the tetrad differences were obtained seem to be due to the influence of overlapping group factors within the tests.

A further analysis of five of the memory tests, which do not seem to duplicate each other, was made to determine whether there is a general factor present in them. The Otis Memory and the Logical Ideas tests were eliminated from this analysis. There are only 15 calculated tetrads from the five variables considered. The median of the obtained tetrads for the intercorrelations between the tests is .0610 and the theoretical probable error calculated from Formula 15 is only .0226; the probable error calculated from Formula 16A is .0216. There are only a limited number of tests involved in this analysis and there are probably many aspects of memory

for verbal material which are not measured, but the obtained median of the tetrad differences is much larger than their theoretical probable error and the intercorrelations between the tests seem to be due to overlapping group factors. These results are significant because the Logical Thought and the Recognition of Syllables tests are both included in the analysis.

## APPENDIX

### TABLE IX

THE INTERCORRELATIONS OF SEVEN MEMORY TESTS WITH TWO TESTS OF INTELLIGENCE AND WITH EACH OTHER, CORRECTED FOR ATTENUATION

| Tests | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Otis Intelligence | — | .990<br>.010 | .295<br>.044 | .274<br>.044 | .479<br>.038 | .812<br>.016 | .717<br>.023 | .777<br>.020 | .245<br>.045 |
| 2. Army Alpha | | — | .201<br>.046 | .249<br>.045 | .469<br>.037 | .832<br>.015 | .685<br>.025 | .721<br>.023 | .222<br>.046 |
| 3. Recognition for Words | | | *.647*<br>*.028* | .671<br>.026 | .259<br>.045 | .358<br>.042 | .436<br>.039 | .418<br>.039 | .352<br>.042 |
| 4. Recognition for Syllables | | | | *.717*<br>*.023* | .281<br>.044 | .349<br>.042 | .498<br>.036 | .450<br>.380 | .162<br>.046 |
| 5. Memory for Dates | | | | | *.762*<br>*.020* | .301<br>.043 | .476<br>.037 | .465<br>.037 | .686<br>.025 |
| 6. Otis Memory Test | | | | | | *.511*<br>*.035* | .618<br>.030 | .633<br>.029 | .237<br>.045 |
| 7. Logical Ideas | | | | | | | *.737*<br>*.022* | 1.103<br>.009 | .448<br>.038 |
| 8. Logical Thought | | | | | | | | *.814*<br>*.016* | .262<br>.044 |
| 9. Recall for Words | | | | | | | | | *.641*<br>*.026* |

VI. *Conclusions:*

1. The Otis Memory Test is neither a reliable nor a valid test of memory. It seems to measure the same mental functions as those measured by the Otis Self-Administering and the Army Alpha intelligence tests. This last conclusion must be tentative, however, because of the unreliability of the test.

2. The analysis by partial correlations indicates that the recognition tests measure memory functions that are related

only to a limited extent to those measured by the other tests. They probably measure pure retentivity uninfluenced by other factors to a greater extent than the more complex tests of meaningful material. Their low correlations with the intelligence tests seem to indicate that there is little relation between pure retentivity and intelligence.

3. The correlations between the logical memory tests and the intelligence tests are very high. The results of this study indicate that the use of the logical memory tests in the Binet intelligence scale does give a good measure of intelligence.

4. It is tentatively concluded from the present study that memory is not a single unitary mental function or trait which is independent of intelligence, but that the intercorrelations between the memory tests are due to overlapping general or group factors in the variables measured by the tests.

5. The reliable residual correlations between some of the memory tests after the intelligence factors measured by the two intelligence tests are controlled indicate that there may be certain memory functions which are largely independent of intelligence. The tests which measure pure retentivity are only slightly related to the intelligence tests.

## References

1. Achilles, E. M., Experimental Studies in Recall and Recognition, *Arch. of Psychol.*, 1920, **44**, 1–80.
2. Binet, A., The Experimental Study of Memory, *Amer. Nat.*, 1897, **31**, 912–916.
3. Binet, A., *Les Idees modernes sur les enfants*, Paris, 1911.
4. Bolton, T. L., The Growth of Memory in School Children, *Amer. J. Psychol.*, 1909, **4**, 362–380.
5. Burt, C., Experimental Tests of General Intelligence, *Brit. J. of Psychol.*, 1909, **3**, 94–177.
6. Carey, N., Factors in the Mental Processes of School Children, *Brit. J. of Psychol.*, 1915–17, **8**, 70–92.
7. Ebbinghaus, H., Über das Gedächtnis, 1885 (translated under title: Memory, a Contribution to Experimental Psychology, 1885, by Ruger, Henry A., and Bussenius, Clara E., Teachers College, Columbia University, 1913). References are to the translation.
8. Garrett, H. E., The Relation of Tests of Memory and Learning to Each Other and to General Intelligence in a Highly Selected Adult Group, *J. Educ. Psychol.*, 1928, **19**, 601–613.

9. GATES, A. I., Correlations of Immediate and Delayed Recall, *J. Educ. Psychol.*, 1918, **9**, 490–496.
10. HOLLINGWORTH, H. L., Characteristic Differences between Recall and Recognition, *Amer. J. of Psychol.*, 1913.
11. HOLZINGER, K., Note on the Use of the Spearman Prophecy-formula for Reliability, *J. Educ. Psychol.*, 1923, **14**, 302–305.
12. HOLZINGER, K., AND CLAYTON, B., Further Experiment on the Applicability of the Spearman Prophecy-formula, *J. Educ. Psychol.*, 1925, **16**, 289–300.
13. HOLZINGER, K., *Statistical Methods for Students in Education*, Ginn and Company, 1928.
14. KELLEY, T. L., *Statistical Method*, The Macmillan Company, 1923.
15. KELLEY, T. L., *Interpretation of Educational Measurements*, World Book Company, 1927.
16. KELLEY, T. L., *Crossroads in the Mind of Man*, Stanford University Press, Stanford University, California, 1928.
17. KIRKPATRICK, E. A., An Experimental Study of Memory, *Psychol. Rev.*, 1894, **1**, 602–209.
18. KITSON, H. D., The Scientific Study of the College Student, *Psychol. Monog.*, 1917, **23**, No. 1.
19. LANIER, L. H., Prediction of the Reliability of Mental Tests and Tests of Special Abilities, *J. Exper. Psychol.*, 1927, **10**, 60–113.
20. LEE, A. L., An Experimental Study of Retention and Its Relation to Intelligence, *Psychol. Monog.*, 1925, **34**, No. 157.
21. MCCALL, W. A., *Correlation of Some Psychological and Educational Measurements*, Teachers College, Columbia University Contributions to Education, No. 79, 1916.
22. OTIS, A. I., *Otis Self-Administering Tests of Mental Ability*, Manual of Directions and Key, World Book Company, 1922.
23. MULHALL, E. F., Experimental Studies in Recall and Recognition, *Amer. J. Psychol.*, 1915, **26**, 217–228.
24. PETERSON, J., Effect of Attitude on Immediate and Delayed Reproduction, *J. Educ. Psychol.*, 1916, **7**, 523–532.
25. PETERSON, J., Frequency and Recency Factors in Maze Learning by White Rats, *J. Animal Behav.*, 1917, **7**, 338–364.
26. PETERSON, J., *Early Conceptions and Tests of Intelligence*, World Book Company, 1925.
27. ROSENOW, C., The Analysis of Mental Functions, *Psychol. Monog.*, 1917, **24**, No. 105.
28. RUGG, H. O., *Statistical Methods Applied to Education*, Houghton Mifflin, 1917.
29. SHARP, S. E., Individual Psychology: A Study in Psychological Method, *Amer. J. Psychol.*, 1899, **10**, No. 3.
30. SIMPSON, B. R., *Correlations of Mental Abilities, Teachers College*, Columbia University Contributions to Education, No. 77, 1912.
31. SPEARMAN, C., The Proof and Measurement of Association Between Two Things, *Amer. J. of Psychol.*, 1904, **15**, 72–101.
32. SPEARMAN, C., The Theory of Two Factors, *Psychol. Rev.*, 1914, **21**, 101–105.
33. SPEARMAN, C., *The Abilities of Man, Their Nature and Measurement*, Macmillan, 1927.

34. TERMAN, L. M., *The Measurement of Intelligence*, Houghton Mifflin, 1916, 142–348.
35. WHIPPLE, G. M., *Manual of Mental and Physical Tests, Part II, Complex Processes*, Warwick and York, 1915, 150–203.
36. WINCH, W. H., Immediate Memory in School Children, *Brit. J. Psychol.*, 1901–02, 1, 127–134.
37. WYATT, S., The Quantitative Investigation of Higher Mental Processes, *Brit. J. Psychol.*, 1913–14, **6**, 109–133.
38. YULE, G. U., *An Introduction to the Theory of Statistics*, Lippincott, 1912.

(Manuscript received April 4, 1930)

# AN EXPERIMENTAL TEST OF THE LAW OF ASSIMILATION

BY K. S. YUM

*University of Chicago*

This investigation is concerned with the following problem: To what extent will a motor response that has been contiguously associated with a given perceptual stimulus be aroused by a similar stimulus with which it has not been associated, and to what extent is the likelihood of arousal a function of the degree and kind of similarity involved?

This problem is never encountered in the early literature of association. Some of the modern writers have assumed the truth of the doctrine, and a few have classified the phenomenon under a separate law of association as the Law of Assimilation[1] by Carr (10), Response by Analogy by Woodworth (66), and the Law of Redintegration by Hollingworth (27.) On the other hand a few like Spencer (1, 54, 59), Bain (1, 59), Ward (58), and Peters (47) have attempted to explain the phenomenon in terms of the conventional Law of Similarity. This law states that the thought of an object tends to arouse the thought of a similar object. In other words, an object makes you think of similar objects that you have known. Here, the similarity involved is between the stimulus and response items. Obviously the law is applicable only to ideational sequences and not to perceptual-motor connections. To account for the fact that a new perceptual object will call out a motor response associated with a similar perceptual stimulus on the basis of this law, they assumed that the new perceptual object first aroused an image or idea of the similar object, and that this image then evoked the motor response in accordance with the Law of Contiguity. In the first connection between the

[1] The term 'Assimilation' was used by Ward (58), Spencer (54), and Drobisch (18), but they meant by the term a particular process of 'blending or fusion' of a presentational continuum—a similarity that amounts to identity only.

new stimulating object and the image or idea of the similar object, the stimulus and response items are similar, and hence this part of the associative sequence is to be classified under the conventional Law of Similarity. Among these writers, the doctrine has been based upon general observation and its truth has never been experimentally tested in a systematic fashion.

Much the same condition is found in the allied field of conditioning responses. We have at present a very considerable amount of experimental work [2] to determine the conditions under which a response will be called out by a substitute stimulus. These experiments in the main, however, have employed the method of contiguous or simultaneous presentation of the two or more stimulating objects. Watson (60) has made a few observations on an infant relevant to the principle of similarity. He first conditioned a fear response to a white rat on the basis of simultaneous presentation, and then made a few tests to discover whether or not this fear response would also be transferred to other objects. He found that the fear response was elicited by a rabbit, to some extent by a dog and a fur coat, slightly by cotton wool, hair of the human head, and a Santa Claus mask, and not at all by the room, the table, and the infant's blocks. He, however, has not discussed in his account the differentiating principle involved, which is presumably that of similarity of the stimulating objects.

The Law of Assimilation is also encountered in the discussions of the development of language, the formation of class concepts, and the process of generalization. Writers frequently comment upon the fact that the naming responses in young children are often elicited by and attached to novel objects on the basis of their apparent similarity to objects with which they are familiar, as when they persist in calling all men 'papa,' or refer to a squirrel as 'Kitty' when it is first seen. Such illustrations are likely to convey the impression that this type of reaction is a relatively infrequent phenomenon—occurring only in these mistaken identifica-

[2] See 11, 12, 13, 42, 60, and others.

tions. Other writers, however, have noted that this transference of the naming responses from familiar to unfamiliar objects on the basis of their similarity occurs automatically all the time, and that this process is of fundamental importance in the development of language and of class concepts. Still other writers have insisted that this process of transference is not limited to the naming responses, but that it is a universal phenomenon applicable to all types of responses. They have also urged that this transference of responses from familiar to novel objects and situations on the basis of similarity which occurs on an automatic and unreflective level with children constitutes the germinal prototype of the process of rational generalization. For example, this general point of view has been explicitly developed by Carr (10). The transference of responses on the basis of similarity has also been recognized as a principle involved in scientific generalization—a fact that is attested by the well-known canon of inductive logic that states that what is true of one situation is also true for all similar situations. While there has been considerable experimental work done in the field of generalization and the formation of class concepts (Moore (40), Fisher (22), Hull (28), Peterson (48), Campion (9), Gengerelli (24), Shaffer (53), Eljasch (19), Hüper (30), and Lohbauer (39)), yet the principle of similarity has been taken for granted and has never been experimentally verified in a systematic way.

This transference of a response from a given stimulating object or situation to a similar stimulus obviously belongs to the general field of transfer. If the doctrine were true, similarity of the stimulating situations would constitute at least one of the conditions under which transfer occurs. Is such similarity one of the conditions of transfer? In spite of the wealth of experimental work on transfer,[3] we know practically nothing as to the exact conditions under which transfer does or does not take place, the conditions under which it is positive or negative, and the factors that determine

[3] Some of the recent articles and a few of the earlier ones of importance are included in the Bibliographical Section (3–8, 14–17, 20, 23, 25, 26, 29, 31–35, 37, 38, 41, 46, 49, 50, 55–57, 61–65).

the amount of transfer. We have a number of theories that purport to state the nature of these conditions, but only one of these employs the concept of similarity—that of Thorndike (55, 57) and Poffenberger (49). But in this theory, the similarity involved is that of bonds of connection, or similarity of the two activities in respect to their stimulus-response constituents, rather than a similarity of the two stimulating situations. Is similarity of the two stimulating situations a favorable or an essential condition of transfer? No one has proposed such a conception, and I know of but two experiments that have attacked the problem from this standpoint —that of Wylie (67) with animal subjects, and that of Bruce (8), an unpublished thesis in the field of memory.

The factual character of the doctrine that responses and attitudes that have been attached to given stimuli are likely to be elicited by similar stimuli with which they have not been so associated in times past, is also involved in the problem of trade mark infringements. Obviously business firms attempt infringement of successful trade marks on the assumption that the public attitudes and responses to these distinctive trade marks will likely be elicited by highly similar ones (Barber (2), Paynter (43–45), and Rogers (51, 52)). Judicial practice, however, seems to hold that an infringement is involved only when the two trade marks are so highly similar as not to be distinguished in ordinary experience. Consequently the various experiments in this field have been primarily concerned with testing the likelihood of memory confusion of a pair of such stimuli (Katzaroff (36), Feingold (21), Paynter (44, 45)), and have not attempted a systematic investigation of the more general principle involved.

The principle at issue is thus involved in such phenomena as association, conditioning, the growth of language ability, the formation of class concepts, generalization, transfer of training, and trade mark infringement. The principle is frequently not even recognized, and when recognized its truth is generally taken for granted. In only a few cases has the principle been subjected to an experimental investigation.

This experiment was initiated on the belief that a principle of such widespread and fundamental importance from the standpoint of the integration and organization of behavior should be experimentally established and systematically investigated. In our procedure the subjects are required to memorize a series of paired associates to the point of mastery. In the recall test, twenty-four hours later, novel stimuli exhibiting certain specified kinds and degrees of similarity, were substituted for the original stimuli. The amount of recall for the substitute stimuli was compared with that for the standard condition in which the original stimuli were employed.

The influence of three kinds of similarity was investigated: 1. Similarity of nonsense syllables in respect to spelling. 2. Similarity of word stimuli in respect to meaning. 3. Similarity of visual patterns.

## Experiment I

### *Similarity of Nonsense Syllables in Respect to Spelling*

The memory materials consisted of seven lists, each containing fourteen paired associates. The stimulus items consisted of two hyphenated nonsense syllables, and the response item was a word of four letters. Each pair was presented for four seconds. All the subjects memorized each of the seven lists and in the same order. Each list was learned in a single sitting, and recall was tested the following day preceding the memorization of the subsequent list. Each list was mastered to the point of a successful immediate recall with a chance order of presentation. All subjects were given one practice list before beginning the experiment proper.

The nature of the stimulus alterations employed in the subsequent recall test are specified and illustrated in Table I. Both the number and the position of the spelling alterations were varied. Each of the seven conditions of recall occurred twice in each list of fourteen paired associates, and the ordinal position of each of these seven conditions was systematically varied from list to list so that each condition occupied every ordinal position.

Records were obtained from thirty-seven subjects. Since each condition occurred twice in each of the seven lists, we obtained from each subject fourteen recall scores, and hence

TABLE I

| Conditions | Original Stimuli | Response Words | Recall Stimuli |
|---|---|---|---|
| Same Syllables | REB–QIM | Wolf | REB–QIM |
| 1st Letter of 1st Syllable Changed | HUD–LEP | Fist | XUD–LEP |
| 2d Letter of 1st Syllable Changed | TOQ–BEX | Jury | TIQ–BEX |
| 1st and 2d Letters of First Syllable Changed | KAJ–ZOY | View | NEJ–ZOY |
| 1st Letter of 2d Syllable Changed | VAH–MIZ | Nose | VAH–PIZ |
| 2d Letter of 2d Syllable Changed | WUL–GIC | Vase | WUL–GOC |
| 1st and 2d Letters of Second Syllable Changed | JEC–POR | Mask | JEC–NAR |

the average group value for each of the seven conditions is based upon 518 scores. The data are given in Table II. The first column lists the seven conditions. The second column gives the average number of response words recalled per individual for the seven lists—the maximum possible

TABLE II

| Conditions | Average No. Recalled | Prob. Err. of Mean | Percentage Recalled |
|---|---|---|---|
| Same Syllables (Control) | 9.4865 | .2992 | 67.761 |
| 1st Letter of 1st Syllable Changed | 5.6216 | .2325 | 40.154 |
| 2d Letter of 1st Syllable Changed | 8.3784 | .2085 | 59.846 |
| 1st and 2d Letters of First Syllable Changed | 5.5135 | .3003 | 39.382 |
| 1st Letter of 2d Syllable Changed | 5.7297 | .2760 | 40.926 |
| 2d Letter of 2d Syllable Changed | 7.5405 | .2241 | 53.861 |
| 1st and 2d Letters of Second Syllable Changed | 5.2703 | .2414 | 37.645 |

value being fourteen. The probable error of these means is also given, and finally the amount recalled is stated in percentage terms.

1. All alterations of the stimulus reduced the amount recalled as compared with the standard condition of no change. All of the differences are statistically significant.

2. An alteration of the first letter of either of the two syllables of the stimulus item reduces the recall more than does a change of the middle or vowel letters of these two syllables. Both differences are statistically significant.

3. A change of two letters in either of the two syllables

reduces the recall score more than does a change in either of the two letters singly. The differences are consistent but they are too small to permit a confident conclusion.

Stated in terms of similarity we may conclude:

1. A substitute stimulus exhibiting a similarity to the original stimulus in respect to four of the six letters elicited a large amount of recall, but an amount that is significantly less than for the standard condition.

2. There is no conclusive evidence that the amount of recall varies significantly with the degree of similarity.

3. The amount of recall varied significantly with the locus of similarity in spelling. Similarity in respect to the initial letter of either syllable is more conducive to recall than a similarity in respect to the middle or vowel components of the two syllables.

## Experiment II

### *Similarity of Word Stimuli in Respect to Meaning*

The memory material consisted of three lists—each containing twelve paired associates. Both stimulus and response items consisted of meaningful words, the response item being a word of six letters. Each pair was presented for three seconds. All the subjects memorized each of the three lists, and in the same order. Each list was learned in a single sitting, and recall was tested on the following day preceding the memorization of the subsequent list. Each list was mastered to the point of a successful immediate recall with a chance order of presentation. All subjects were given one practice list before beginning the experiment proper.

In the subsequent recall, twenty-four hours later, three conditions were employed. In one condition the original stimulus word was retained as a norm for comparison. In the other two conditions, new words were substituted for the stimulus words, and these new words exhibited one of two degrees of meaning similarity to the original stimulus words.

The procedure of selecting these similar words was as follows: For each stimulus word of the lists, the experimenter assembled a series of words exhibiting a wide range of similarity in respect to meaning. These words were then ranked

by twenty judges in respect to their degree of similarity to the original stimulus words. From each series the experimenter selected for the recall test those words which exhibited two degrees of similarity according to the unanimous opinion of the judges.

Samples of the original stimulus words, the two similar words, and the response words are given in Table III. The

TABLE III

THE STIMULUS WORDS USED IN RECALL TEST ARE ITALICIZED

| Original Stimulus Words | Degree of Similarity | | Response Words |
|---|---|---|---|
| | 1st Degree | 2d Degree | |
| SNAKE | *SERPENT* | TURTLE | WEALTH |
| *FIGHT* | BATTLE | ARGUMENT | LEDGER |
| HOUSE | COTTAGE | *BARN* | BREEZE |

words that were employed as stimuli in the subsequent recall test are italicized. Each of the three conditions of recall occurred four times in each list of twelve paired associates, and the ordinal position of each of these three conditions was systematically varied from list to list so that each condition occupied every ordinal position.

Records were obtained from fifty-four subjects. Since each condition occurred four times in each of the three lists, we obtained from each subject twelve recall scores, and hence the average group value for each of the three conditions is based on 648 scores. The data are given in Table IV. The

TABLE IV

| Conditions | Arithmetic Mean | Prob. Err. of Mean | Percentage Recalled |
|---|---|---|---|
| Same Word (Control) | 6.0185 | .2167 | 50.154 |
| 1st Degree of Similarity | 3.9074 | .2165 | 32.562 |
| 2d Degree of Similarity | 1.3519 | .1640 | 11.266 |

first column lists the three conditions. The second column gives the average number of response words recalled per individual for the three lists—the maximum possible value

being twelve. The probable errors for these means are also given, and finally the amount recalled is stated in percentage terms. The differences between the means are thoroughly reliable, and the mean values for the individual subjects exhibit the same trend with but three inversions of order in the one hundred and eight comparisons.

The results prove quite conclusively that a response associated with a stimulus word tends to be aroused by a novel word having a similar meaning, and that the likelihood of arousal varies directly with the degree of similarity.

## Experiment III

### *Similarity of Visual Patterns*

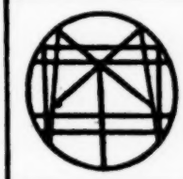
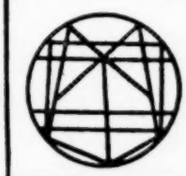
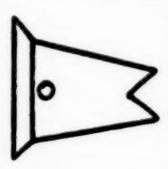
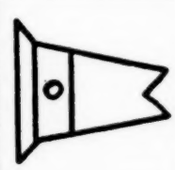
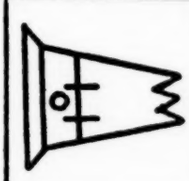
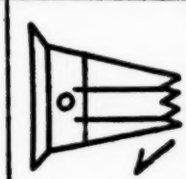
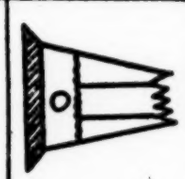
The memory material consisted of three lists—each containing fifteen pairs of items. Each pair consisted of a visual pattern as the stimulus and a word of three letters as the response item, and it was presented for three seconds. All the subjects memorized each of the three lists, and in the same order. Each list was learned in a single sitting, and recall was tested on the following day preceding the memorization of the subsequent list. Each list was mastered to the point of a successful immediate recall with a chance order of presentation. All subjects were given one practice list before beginning the experiment proper.

Five conditions of recall were employed in the subsequent test, twenty-four hours later. In the first condition the original visual pattern was retained as a norm for comparison. In the other four conditions, new visual patterns were substituted for the original stimuli, and they exhibited one of four degrees of similarity to those stimuli. The visual patterns employed as stimuli were drawn by the experimenter, and for each of the original stimuli were drawn a series of patterns exhibiting a wide range of similarity of appearance. Each series of patterns was arranged in order of similarity by each of fifteen judges, and from each series were selected for the recall test four patterns exhibiting four gradations of similarity according to the unanimous opinion of the judges.

Samples of the patterns exhibiting all four gradations of similarity are given in Table V. Those patterns used in the

TABLE V

THE STIMULI USED IN RECALL TEST ARE CHECKED

| ORIGINAL STIMULI | DEGREE OF SIMILARITY | | | |
|---|---|---|---|---|
| | IST. DEGREE | 2ND. DEGREE | 3RD. DEGREE | 4TH. DEGREE |
| | | |  |  |
| | | | | |
| | | | | |
|  |  |  |  |  |
| | | | | |

recall test are checked. Each of the five conditions of recall occurred three times in each list of fifteen paired associates, and the ordinal position of each of these five conditions was systematically varied from list to list.

Records were obtained from twenty-six subjects. Since each condition occurred three times in each of the three lists, we obtained from each subject nine recall scores, and hence the average group value for each of the five conditions is based upon 234 scores. The data are given in Table VI.

The first column lists the five conditions. The second column gives the average number of response words recalled per individual for the three lists—the maximum possible value being nine. The probable errors for these means are

TABLE VI

| Conditions | Average No. Recalled | Prob. Err. of Mean | Percentage Recalled |
|---|---|---|---|
| Same Pattern (Control) | 7.6154 | .1684 | 84.616 |
| First Degree of Similarity | 5.8077 | .2349 | 64.530 |
| Second Degree of Similarity | 4.4231 | .2144 | 49.146 |
| Third Degree of Similarity | 4.0769 | .2229 | 45.299 |
| Fourth Degree of Similarity | 3.2692 | .2218 | 36.324 |

also given, and finally the amount recalled is stated in percentage terms.

The average amount recalled varies directly with the degree of similarity without a single exception. The difference between the normal condition and the first degree of similarity is statistically reliable, and the averages for all of the subjects conform to the group results. The difference between the first and the second degree of similarity is also reliable and only three of the individual means fail to conform to the average results for the group. The difference between the second and the third degree of similarity is not statistically reliable, and the same order of difference is exhibited by ten of the twenty-six subjects, and the reverse order by four subjects of the group. The difference between the third and the fourth degree of similarity does not exhibit any high degree of reliability, but the same order of difference is exhibited by sixteen subjects and the reverse order by but two subjects of the group of twenty-six individuals.

We may thus conclude that responses associated with a visual pattern tend to be aroused by a similar visual pattern, and that the likelihood of recall tends to vary directly with the degree of similarity.

## Conclusions

1. Our experiment has established the fact that a naming response that is contiguously associated with a visual stimulus

is likely to be aroused by a visual stimulus with which it has not been so associated, whenever the two stimuli exhibit certain kinds and degrees of similarity.

2. The effectiveness of three kinds of similarity has been demonstrated—similarity of nonsense syllables in respect to spelling, similarity of word stimuli in respect to meaning, and similarity of visual patterns. There are no data on the relative effectiveness of these three kinds of similarity.

3. With similarities of meaning and visual patterns, the likelihood of recall varied directly with the degree of similarity. There was no evidence that the likelihood of recall varies with the amount of similarity of nonsense syllables as measured by the number of letters in common.

4. With nonsense syllables employed as stimuli, the likelihood of recall varied with the locus of the common letters.

5. Presumably this phenomenon can be obtained with other modes of stimulation, auditory, *e.g.*, with other than naming responses, with other kinds of similarity, and with wider degrees of dissimilarity than those employed. It is presumed that we are here dealing with a general principle, whose limitations are as yet undefined.

## Bibliography


1. Arnold, Felix, The Psychology of Association, *Arch. Psychol.*, 1906, No. 3, N. Y., Columbia Univ.
2. Barber, A. W., The Psychology of Trade-Marks, *Bull. U. S. Trade-Mark Assn.*, 1911, 152–165.
3. Bode, B. H., *The Reaction against Formal Discipline.* Conflicting Psychologies of Learning, 1929, N. Y., Heath.
4. Bray, C. W., Transfer of Learning, *J. Exp. Psychol.*, 1928, 11, 443–467.
5. Brolyer, C. R., Thorndike, E. L., and Woodyard, E., A Second Study of Mental Discipline in High School Studies, *J. Educ. Psychol.*, 1927, 18, 377–404.
6. Brooks, F. D., The Transfer of Training in Relation to Intelligence, *J. Educ. Psychol.*, 1924, 15, 413–422.
7. Brown, W., Habit Interference in Sorting Cards, *Univ. of Cal. Pub. in Psychol.*, 1914, 1, 269–321.
8. Bruce, R. W., *Conditions of Transfer of Training*, 1929, Univ. of Chicago, Ph.D. thesis.
9. Campion, G. G., The Organic Growth of the Concept as One of the Factors in Intelligence, *Brit. J. Psychol.*, 1928, 19, 60–64.
10. Carr, H. A., *Psychology*, A Study of Mental Activity, 1925, N. Y., Longmans, Green & Co., 99–106; 212–214; 237–239; 348–351.

11. Cason, H., The Conditioned Pupillary Reaction, *J. Exp. Psychol.*, 1922, **5**, 108–146.
12. Cason, H., The Conditioned Eyelid Reaction, *J. Exp. Psychol.*, 1922, **5**, 153–196.
13. Cason, H., The Conditioned Reflex or Conditioned Response as a Common Activity of Living Organisms, *Psychol. Bull.*, 1925, **22**, 445–464.
14. Cole, L. W., Interference between Related Mental Processes, *J. Educ. Res.*, 1928, **18**, 32–39.
15. Coover, J. E., Formal Discipline from the Standpoint of Experimental Psychology, *Psychol. Monog.*, 1916, 20.
16. Dashiell, J. F., Some Transfer Factors in Maze Learning by the White Rats, *Psychobiol.*, 1920, **2**, 329–350.
17. De Weerdt, E. H., The Transfer Effect of Practice in Related Functions upon a Group Intelligence Test, *Sch. and Soc.*, 1927, **25**, 438–440.
18. Drobisch, M. W., *Empirische Psychologie*, Erste Auflage, 1842, 142.
19. Eljasch, M., Neue Abstraktionsversuche bei Vorschulpflichtigen Kindern, *Zsch. f. Psychol.*, 1927, **105**, 1–42.
20. Ewert, P. H., Bilateral Transfer in Mirrow-Drawing, *Ped. Sem.*, 1926, **33**, 235–249.
21. Feingold, G. A., Recognition and Discrimination, *Psychol. Monog.*, 1915, 18.
22. Fisher, S. C., The Process of Generalizing Abstraction and its Product, the General Concept, *Psychol. Monog.*, 1916, 21.
23. Gates, A. I., A Critique of Methods of Estimating and Measuring the Transfer of Training, *J. Educ. Psychol.*, 1924, **15**, 545–558.
24. Gengerelli, J. H., Mutual Interference in the Evolution of Concepts, *Amer. J. Psychol.*, 1927, **38**, 639–646.
25. Hadley, C. V. D., Transfer Experiments with Guinea Pigs, *Brit. J. Psychol.*, 1927, **18**, 189–224.
26. Ho, Y. H., Transfer and Degree of Integration, *J. Comp. Psychol.*, 1928, **8**, 87–99.
27. Hollingworth, H. L., Psychology, its Facts and Principles, 1928, N. Y., Appleton.
28. Hull, C. L., Quantitative Aspects of the Evolution of Concepts, *Psychol. Monog.*, 1920, 28.
29. Hunter, W. S., The Interference of Auditory Habits in the White Rat, *J. Animal Behav.*, 1917, **7**, 49–65.
30. Hüper, H., Über die Vermendung der Achschen Suchmethode zur Analyse der Begriffsbildung. Eine Experimentelle und Kritische Untersuchung, *Arch. f. d. ges. Psychol.*, 1928, **62**, 315–408.
31. Jablonower, J., Opportunities for Transfer Making in Elementary Mathematics, *Teachers Coll. Rec.*, 1929, **30**, 808–818.
32. Judd, C. H., Practice and its Effects on the Perception of Illusions, *Psychol. Rev.*, 1902, **9**, 27–39.
33. Judd, C. H., The Relation of Special Training and General Intelligence, *Educ. Rev.*, 1908, **36**, 28–42.
34. Judd, C. H., *Generalized Experience.* Psychology of Secondary Education, 1927, Chicago, Ginn, Chap. 19.
35. Judd, C. H., and Cowling, D. J., Studies in Perceptual Development, *Psychol. Monog.*, 1907, **7**, 349–369.
36. Katzaroff, D., La Recognition, *Arch. de Psychol.*, 1911, **11**, 1–78.
37. Kline, L. W., Some Experimental Evidence in Regard to Formal Discipline, *J. Educ. Psychol.*, 1914, **5**, 259–266.
38. Langdon, J. N., and Yates, E. M., An Experimental Investigation into Transfer of Training in Skilled Performance, *Brit. J. Psychol.*, 1928, **18**, 422–437.

39. LOHBAUER, H., Über die Begriffe von Psychischen im Kindesalter, *Arch. f. d. ges. Psychol.*, 1928, **64**, 133–214.
40. MOORE, T. V., The Process of Abstraction, *Univ. of Calif. Pub. in Psychol.*, 1910, **1**, 73–197.
41. ORATA, P. T., *The Theory of Identical Elements*, 1928, Columbus, The Ohio State Univ.
42, PAVLOV, I. P., *Conditioned Reflexes*, 1927, London, Oxford Univ.
43. PAYNTER, R. H., Experiment versus Court Decision, *J. Philos. Psychol., etc.*, 1915, **12**, 45–47.
44. PAYNTER, R. H., A Psychological Investigation of the Likelihood of Confusion between the Words 'Coca-Cola' and 'Chero-Cola,' *J. Appl. Psychol.*, 1919, **3**, 329–351.
45. PAYNTER, R. H., A Psychological Study of Trade-Mark Infringement, *Arch. Psychol.*, 1920, **6**, No. 42.
46. PEARCE, B. D., A Note on the Interference of Visual Habits in the White Rats, *J. Animal Beh.*, 1917, **7**, 169–177.
47. PETERS, W., Über Ähnlichkeitsassoziation, *Zsch. f. Psychol.*, 1910, **56**, 161–206.
48. PETERSON, J. C., The Higher Mental Processes in Learning, *Psychol. Monog.*, 1920, 20.
49. POFFENBERGER, A. T., JR., The Influence of Improvement in One Simple Mental Process upon other Related Processes, *J. Educ. Psychol.*, 1915, **6**, 459–474.
50. PYLE, W. H., Transfer and Interference in Card-Distributing, *J. Educ. Psychol.*, 1919, **10**, 107–110.
51. ROGERS, E. S., The Unwary Purchaser, a Study in the Psychology of Trade-Mark Infringement, *Mich. Law Rev.*, 1910, **8**, 613–622.
52. ROGERS, E. S., *Good Will Trade-Marks, and Unfair Trading*, 1914.
53. SHAFFER, L. F., The Measurement of Children's Concepts, *J. Educ. Psychol.*, 1928, **18**, 41–44.
54. SPENCER, H., *Principles of Psychology*, 1892, 3d Ed., 1, N. Y., Appleton, Chaps. 7 and 8, 250–271.
55. THORNDIKE, ED. L., *Educational Psychology*, Briefer Course, 1914, N. Y., Teachers College, 268–269.
56. THORNDIKE, ED. L., Mental Discipline in High-School Studies, *J. Educ. Psychol.*, 1924, **15**, 1–22; 83–98.
57. THORNDIKE, ED. L., AND WOODWORTH, R. S., The Influence of Improvement in One Mental Function upon the Efficiency of Other Functions, *Psychol. Rev.*, 1901, **8**, 247–261; 384–395; 553–564.
58. WARD, J., *Psychological Principles*, 1919, N. Y., Putnam's Sons, 82ff.; 191–192.
59. WARREN, H. C., *A History of the Association Psychology*, 1921, Charles Scribner's Sons.
60. WATSON, J. B., *Psychology from the Standpoint of a Behaviorist*, 1924, Phila., Lippincott, 2d Ed., 233–235.
61. WEBB, L. W., Transfer of Training and Retroaction, *Psychol. Monog.*, 1917, 24.
62. WHIPPLE, G. M., The Transfer of Training, *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 2, 1928, 179–209.
63. WILTBANK, R. T., Transfer of Training in White Rats upon Various Series of Mazes, *Behav. Monog.*, 1919, **4**, 1–65.
64. WINCH, W. H., The Transfer of Improvement in Reasoning in School-Children, *Brit. J. Psychol.*, 1923, **13**, 370–381.

65. WOODROW, H., The Effect of Types of Training upon Transference, *J. Educ. Psychol.*, 1927, **18**, 159–172.
66. WOODWORTH, R. S., *Psychology*, a Study of Mental Life, 1921, N. Y., Holt, 406–408; 434–435; 452–455. Revised Edition, 1929, 458ff.
67. WYLIE, H. H., An Experimental Study of Transfer of Response in the White Rat, *Behav. Monog.*, 1919, 3.

(Manuscript received April 1, 1930)

# THE FACTORS OF SPEED AND POWER IN TESTS OF INTELLIGENCE

BY FRANK S. FREEMAN

*Cornell University*

The rôles played by speed and power as factors in intelligence have been examined both speculatively and experimentally. The results of these investigations have not been uniform, due in part, probably, to the fact that there has not been, nor is there at present, agreement on just what constitutes intelligence. Lack of agreement regarding the very function we seek to understand must necessarily lead to differently conceived experiments and to a variety of results and conclusions. The author of a recent paper [1] enumerates briefly some of the experiments performed over a span of more than a half-century. These include lifting of weights, discrimination of pitch and clang, rate and accuracy of computation, rate of reading and retention of ideas, thrusting at targets, maze tracing, drawing straight lines, and rate of typewriting. This same author himself studied the interrelations of speed, accuracy and difficulty, as revealed in the spelling of words ranging from the very easy to the very difficult. He concludes, among other things, that the fast speller on the basis of general average will probably rank fairly high in accuracy of spelling.

Peak and Boring [2] report the results of an investigation in which the Army Alpha and the Otis Self-Administering Test of Mental Ability were used with five advanced students of psychology. They conclude that there is a high correlation between score in an intelligence test, speed in the test, and speed in a simple reaction, and, furthermore, that speed of reaction is the most important factor in individual differences in the intelligent act.

[1] L. S. McLeod, The Interrelations of Speed, Accuracy and Difficulty, *J. Exp. Psy.*, Vol. 12, No. 5, pp. 431–443.

[2] H. Peak and E. G. Boring, The Factor of Speed in Intelligence, *J. Exp. Psy.*, 1926, Vol. 9, No. 2, pp. 71–94.

The diversity of the experimental work is apparent; and the inadequacy of present results for purposes of widespread and general application should likewise be apparent. Rate and accuracy in reading, in spelling, in weight lifting, etc., may very well be significant in and for themselves, but the application of these results to other situations is of doubtful validity. We may ask, for example, what is the relationship of rate to the quality of creative effort, as, for example, in musical composition, the writing of prose or poetry, scientific experimentation, painting and the like? How are rate and validity of critical evaluation related? What is the relationship between speed of decision and excellence or poorness of results in business enterprises? Further, what is the relationship between the skilled artisan's rate of work and the quality of his product? These questions—particularly the first—are extremely important in that type of activity which we are pleased to call intelligent. But to these no answer can be given from the results of previous studies; nor, it should be added hastily, will one be supplied in this study.

In the present instance, however, the writer is concerned with the answer to a very legitimate question: namely, do our current tests of intelligence examine speed or power in the solution of the situations presented by them? Are the aforementioned conclusions of Peak and Boring correct? It should be clear from our results that we are not attempting to generalize with respect to all intelligent behavior. The degree to which the present study can be elsewhere applied will depend upon the extent of common ground found between the tests and other situations.

This same problem was considered by the writer in an earlier paper [3] in which it was pointed out that intelligence as measured by the Dearborn Group Test was chiefly a matter of power, with speed playing an insignificant rôle. Since it is known, however, that not all tests measure exactly the same mental processes, the conclusions based on one cannot readily be applied to others. Therefore, in order to determine how far the tests agree or disagree, the investigation

[3] F. S. Freeman, Power and Speed: Their Influence upon Intelligence Test Scores, *J. of App. Psych.*, Vol. 12, No. 6, pp. 631–635.

here reported was undertaken, using the National Intelligence Test (Scale A, Form 1), the Otis Advanced Examination (Form A), and the Terman Group Test of Mental Ability (Form A). It will be recalled that the first of these is intended for grades 3–8, the second for 5–12 and the third for 7–12. The National was used with 56 pupils in grade 6, the Otis with 51 in grade 8, and the Terman with 42 in grade 11. These particular tests were selected because they are among the most widely employed.

Each test was first administered with the regular time limit. The next day the same test was given, but the subjects were permitted to spend as much time as they needed. In fact, they were instructed to attempt every item and to work with great care inasmuch as there would be time enough for all.

In the writer's previous paper (p. 631) it was stated that earlier investigations, in which time limits were simply doubled, yielded ambiguous results because the time factor had not been properly controlled, for "if a test is purely a test of rate, then the correlation [for regular time and double time scores] will be high *if* the items are sufficient in number to permit few or none of the subjects to complete them all within the time limit. It is clear that if the time allowance is of such length that everyone has an opportunity to attempt every item, time ceases to be a factor and power only is significant. However, the converse—that a high correlation [when the time limits are doubled] means the test is a measure of speed alone, or chiefly—is doubtful; in fact, highly improbable.

"If the time is doubled, but if all the subjects are not enabled thereby to attempt every item, the correlations for single and double time will still be ambiguous, for both speed and power will have entered into the results achieved." The time factor should, therefore, be eliminated in the retesting if we wish to examine the contributions of power and speed to performance. By so doing we are enabled to derive correlations for scores achieved with a time limit imposed and for scores made without a time limit, and,

therefore, not influenced by rate of work. "If under these conditions there is a high correlation, then we may justifiably conclude that the test is primarily one of power. . . . If, however, there is a low correlation, we may conclude that the test with its original time limit measures speed for the most part, inasmuch as it appears that the slow, plodding, accurate individual is able to score much higher if he is given sufficient time to undertake the solution of each problem," [4] whereas the subjects whose principal quality is rapidity of response may profit little or none from the increase in time for they "will reach an impasse beyond which progress will not be possible even with the added opportunity of attempting the solutions of the increasingly difficult items." [5]

With this view in mind, the following tables may be examined. It will be noted that the total scores of the National Test yield a rather high correlation coefficient. According to our hypothesis, therefore, the test as a whole would be characterized as one measuring power, principally,

TABLE I

NATIONAL INTELLIGENCE TEST

*Correlation of Scores Obtained under Regular Time Limit and Scores Obtained without a Time Limit*

| | $r$ |
|---|---|
| Test I | $.66 \pm .05$ |
| II | $.84 \pm .02$ |
| III | $.55 \pm .06$ |
| IV | $.70 \pm .04$ |
| V | $.19 \pm .09$ |
| Total scores | $.83 \pm .02$ |

(N = 56)

rather than speed. The individual parts, however (with the exception of II, Sentence Completion), show markedly lower correlations, notably number V, where the relationship is altogether negligible: namely, .19. Inspection of Part V will readily indicate that this condition is exactly what is to be expected, for it is a Symbol-Digit test, in which the subject simply matches various symbols and digits according to a given key to which he may refer at any time. Thus Part V is distinctly a test of speed of performance; and by comparing

[4] *Op. cit.*, p. 632.

[5] *Op. cit.*, p. 632.

the several coefficients the relative importance of speed in each of the other parts may be inferred.

Yet, though the National Test be one of power in the main, it is important to note that the correlation is certainly not unity. The coefficient is far enough removed from unity to indicate that some individuals are handicapped through the imposition of a time-limit, as revealed by the more-than-chance shifting in rank, which, of course, will result in lowering the correlational index. In other words, there are those, relatively few in number, who might be regarded as 'slow but accurate.'

TABLE II

OTIS ADVANCED EXAMINATION

*Correlation of Scores Obtained under Regular Time Limit and Scores Obtained without a Time Limit*

| | $r$ |
|---|---|
| Test I | .42 ± .08 |
| II | .59 ± .06 |
| III | .18 ± .09 |
| IV | .67 ± .05 |
| V | .54 ± .06 |
| VI | .29 ± .08 |
| VII | .56 ± .06 |
| VIII | .60 ± .06 |
| IX | .47 ± .07 |
| X | .25 ± .08 |
| Total scores | .58 ± .06 |

(N = 51)

From the correlations obtained with the Otis Advanced Examination, we must conclude, if our hypothesis be correct, that it is a test which measures speed of performance to a much greater extent than the Dearborn,[6] the National or the Terman (see Table III and below). The correlation coefficient for the total scores (.58 ± .06) is only moderately high, and so far removed from unity as to indicate a decided shifting in the ranks of the subjects.

As has been stated, the greater the factor of speed, the

[6] F. S. Freeman, *op. cit.*, p. 633.

*Dearborn Correlations*

| | $r$ |
|---|---|
| Part I | .81 ± .01 |
| II | .76 ± .02 |
| III | .64 ± .02 |
| IV | .91 ± .01 |
| Total scores | .88 ± .01 |

smaller will be the degree of correlation between scores achieved, first under a time limit, and then without a time limit. Thus the speed factor is particularly marked in some of the individual parts, notably III, VI, and X, and, to a lesser degree in I and IX. Throughout, however, for both totals and part scores, the correlations are such as to warrant the conclusion that the Otis Advanced Examination, as it stands, emphasizes the factor of speed considerably more than the other tests included in this study and in the earlier one with the Dearborn Test.

In Table III we find a very high correlation existing

TABLE III

TERMAN GROUP TEST OF MENTAL ABILITY

*Correlation of Scores Obtained under Regular Time Limit and Scores Obtained without a Time Limit*

| | r |
|---|---|
| Test I | .82 ± .03 |
| II | .60 ± .06 |
| III | .75 ± .04 |
| IV | .73 ± .05 |
| V | .63 ± .06 |
| VI | .82 ± .03 |
| VII | .70 ± .05 |
| VIII | .64 ± .06 |
| IX | .71 ± .05 |
| X | .87 ± .02 |
| Total scores | .93 ± .01 |

(N = 42)

between the total scores; so high, in fact, as to indicate that the speed factor is almost negligible in the determination of the score on the Terman Group Test. Although this is true of the total score, it does not apply to each of the parts taken independently, for it will be observed that the coefficients range from .60 to .87. Yet for the separate parts, three of the coefficients are moderately high, four high, and three very high. Taking the Terman Test as a whole, therefore, it may be said that as at present constituted it is very largely a measure of power.

In the case of no one of the four intelligence tests (Dearborn, National, Otis, and Terman) can it be maintained that speed is being measured to an appreciable extent for most individuals; but our results do indicate that their demands

on the functions speed and power are not entirely uniform for all persons, and that the Otis seems to measure rate of work more than any of the others, although even it cannot be designated as merely a test of rate of response.

It is important, however, to recognize that speed does play a rôle—even if relatively unimportant for most subjects —in these tests. This is shown not only by the fact that the correlation coefficients are removed from unity, but also by the data in Table IV below.

TABLE IV

COEFFICIENTS OF VARIABILITY[7]

(*For total scores*)

| | *Time Limit* (x) | *No Time Limit* (y) |
|---|---|---|
| National | 15.3 | 12.2 |
| Otis | 12.4 | 10.3 |
| Terman | 18.4 | 14.5 |

It wi l be observed that, in each instance above, the variability under no time-limit is smaller than when the time limit is imposed. In the case of the National, the coefficient of variability under (x) is 25.4 percent greater than under (y); for the Otis it is 20.4 percent greater; and for the Terman 20.7 percent greater. These differences indicate, of course, a decrease in the variability of the distributions under (y). The writer interprets this decrease to mean that with the time handicap removed some subjects will improve their scores relative to the group, whereas others who might be more rapid but less accurate in their responses will lose the peculiar advantage with which a time limit favors them.

We cannot agree with Peak and Boring when they declare that "a great lengthening of time limits for a standard test destroys these correlations [between score and speed in an intelligence test] . . . because the faster subjects do not have an opportunity to utilize the additional time allowed."[8] This would be true if the tests were measures very largely of speed, as stated earlier in this paper. But of the four corre-

[7] C of V $= 100 \frac{\text{S.D.}}{\text{M}}$

[8] *Op. cit.*, p. 92.

lations (Dearborn .88, National .83, Terman .93 and Otis .58),[9] only one is anything but very high. Nor can we, therefore, agree that "speed of reaction is an important, and probably *the most important*,[10] factor in individual differences int he intelligent act." Speed of reaction is of some importance in the Otis Test but not of marked significance in the others. If the view of Peak and Boring were correct we should expect coefficients of correlation at or very close to zero, instead of the .58, .83, .88 and .93 which we have found. On the contrary, therefore, in view of our results, we must conclude that for the large majority of subjects speed of reaction is of very minor importance in tests of intelligence; bearing in mind, however, that there is a small but ever present group who are handicapped by time limits in attempting to achieve a score which shall correctly represent their levels of ability.

## Conclusions

1. The four tests considered in this study are measures chiefly of power, while the factor of speed is of relatively little importance. The only possible exception to this is the Otis Advanced Examination, in which the factor of speed is of fair importance.

2. There is a small but ever-present group of individuals who are penalized by an imposed time-limit and whose test-scores, therefore, are not truly representative of the relative levels they are capable of attaining. This group, though small, presents a genuine problem in the interpretation of the score of any given individual.

3. We must deny the conclusions of Peak and Boring with respect to the importance of speed in determining individual differences in the intelligent act.

4. If the tests considered in this study are representative, then these conclusions may be drawn for current tests in general. It is the writer's belief, however, that each test should be subjected to the same experiment.

[9] F. S. Freeman, *op. cit.*, p. 633.

[10] Italics mine.

(Manuscript received January 13, 1930)

# THE VARIABILITY OF AFFECTIVE JUDGMENTS UPON ODORS

BY J. G. BEEBE-CENTER

*Harvard University*

The present note is a supplement to the articles of P. T. Young [1] and J. B. Kenneth [2] upon the constancy of affective judgments to odors, and also to a discussion of this topic contained in an article upon the recollection of odors by Kate Gordon.[3] The work of these investigators yields information concerning two forms of variability in affective judgments upon odors, namely variability in the judgments of different observers and variability in the judgments of single observers at different times. It does not, however, involve comparison of these two forms of variability. Such comparison is made possible by the data presented below.

In the course of an investigation of affective habituation carried out at Harvard in 1925–6 I determined the affective rank orders of 14 olfactory substances for each of 8 observers in each of three experimental series. The first experimental series (Series I) was separated from the second (Series II) by two weeks, and the second from the third (Series III) by four and one-half to seven months. During intervals between the three series the observers participated in experiments upon affective habituation involving presentation of a large number of olfactory substances including the 14 referred to above.

In Table I will be found, for each of Series I, II and III, the average, range and standard deviation of the twenty-eight coefficients of correlation secured by intercorrelating the rank

[1] P. T. Young, Constancy of Affective Judgments to Odors, J. Exper. Psychol., 1923, 6, 182.

[2] J. H. Kenneth, A Few Odor Preferences and their Constancy, J. Exper. Psychol., 1928, 11, 56.

[3] K. Gordon, The Recollection of Pleasant and Unpleasant Odors, J. Exper. Psychol., 1925, 8, 230.

orders of the 14 stimuli for the 8 observers. The table also shows the average of the averages of the coefficients of

TABLE I

CORRESPONDENCE OF RANK ORDERS FOR DIFFERENT OBSERVERS

| | Average | Range | S.D. |
|---|---|---|---|
| Series I | .54 | −.21 to .88 | .21 |
| Series II | .33 | −.43 to .86 | .33 |
| Series III | .39 | −.16 to .81 | .26 |
| Avg. of Avgs. | .42 | Tot. Rge. −.43 to .88 | |

correlation for all three series, and the total range of these coefficients.

In Table II will be found, for each pair of series, the

TABLE II

CORRESPONDENCE OF RANK ORDERS FOR SAME OBSERVERS ON DIFFERENT OCCASIONS

| | Average | Range | S.D. |
|---|---|---|---|
| Ser. I and II | .83 | .69 to .96 | .09 |
| Ser. II and III | .83 | .58 to .99 | .13 |
| Ser. I and III | .75 | .50 to .94 | .14 |
| Avg. of Avgs. | .80 | Tot. Rge. .50 to .99 | |

average, range and standard deviation of the coefficients of correlation between the rank orders of the 14 stimuli as determined for each observer in the two series. The table also shows the average of the averages of the coefficients of correlation for the three pairs of series and the total range of these coefficients.

Comparison of these two tables shows clearly [4] that for intervals up to seven and one-half months the correspondence between affective rank orders of a single set of olfactory stimuli established at different times for a single observer is much greater than is the correspondence between affective rank orders of such a set of stimuli established for different observers.

[4] The significance of the differences involved is so obvious from mere inspection that statistical proof of it has been omitted.

The tables also yield information concerning the parts played by 'nature' and 'nurture' in relative affective judgments upon odors. Table II shows that the correspondence between rank orders established at different times for the same observer is greater between Series I and II and between Series II and III than it is between Series I and III. The differences involved have a certain statistical significance, being respectively 2 times and 1.7 times their P.E.[5] As the temporal order of the series was I, II and III, this result indicates a gradual shift in the relative affective judgments of an observer as a function of repeated stimulation.[6] Table I shows that the correspondence between rank orders for different observers is greatest for Series I. The differences involved have considerable statistical significance, that between Series I and II being 4 times its P.E., and that between Series I and III 3 times its P.E. This result indicates that the gradual shift mentioned above is in the direction not of greater agreement among different observers, but of greater disagreement. Although these results need confirmation— especially those in Table II— it seems reasonable to conclude tentatively that in the case of relative affective judgments by different individuals upon a single set of olfactory stimuli agreement is primarily dependent upon like organic constitution whereas disagreement rests in the main upon unlike past experiences.[7]

[5] The averages being compared are averages of only eight coefficients. The P.E.'s of their differences are consequently of low reliability.

[6] There is slight evidence of a similar gradual shift in the relative affective value of colors. *Cf.* E. J. G. Bradford, A Note on the Relation and œsthetic Value of the Perceptive Types in Color Appreciation, *Amer. J. Psychol.*, 1913, **24**, 545.

[7] By means of Spearman's statistical technique I recently demonstrated that the relative affective judgments upon any one stimulus in Series I of the experiment described above could be considered to involve a 'general factor' common to the judgments of all eight observers and 'special factors' peculiar to the judgment of each individual observer. (*Cf.* J. G. Beebe-Center, General Affective Value, *Psychol. Rev.*, 1929, **36**, 472.) The conclusion above obviously implies that such 'general factors' depend primarily upon the like organic constitution of the observers while the 'special factors' depend primarily upon their different past experience.

(Manuscript received January 17, 1930)